

UNIVERSIDAD CARLOS III DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

DEPARTAMENTO DE

TEORÍA DE LA SEÑAL Y COMUNICACIONES



GRADO EN INGENIERÍA

DE SISTEMAS DE COMUNICACIONES

TRABAJO DE FIN DE GRADO

SIMULACIÓN DE PÉRDIDAS DE AUDICIÓN Y MEDIDAS
DE INTELIGIBILIDAD

AUTORA: PATRICIA REDONDO MARTÍNEZ

TUTORA: CARMEN PELÁEZ MORENO

Agradecimientos

En primer lugar quiero agradecer a mi tutora *Carmen Peláez Moreno*, por darme la oportunidad de realizar este trabajo de Fin de Grado y por toda su dedicación semana tras semana ayudándome y animándome a la realización de este proyecto.

A mis padres, *Ana Isabel y Tomás*, por el apoyo recibido durante todos estos años. Gracias por creer en mí siempre y animarme en la distancia. A mi hermana mayor, *Laura*, por demostrarme que todo esfuerzo tiene su recompensa, y que retrasar una victoria no significa perder.

Por último, a mis eternos compañeros, *Alex, Elena, Nuria y Sandra*, os aseguro que sin vuestra compañía, tanto en los meses duros de estudio como celebrando los aprobados, estos años no hubieran sido igual.

A todos vosotros, **GRACIAS**

Índice general

ABSTRACT	1
1. Introduction and objectives	1
2. Development of the project	1
2.1 Theory	1
3. Experiment	3
4. Planification	7
5. Conclusion	7
1. INTRODUCCIÓN Y OBJETIVOS	8
1.1 Motivaciones	8
1.2 Planteamiento del problema	9
1.3 Objetivos	10
1.4 Estructura del proyecto	11
2. CONCEPTOS PRELIMINARES	13
2.1 Análisis del Estado del Arte	13
2.2 El sonido	14
2.3 La voz humana	15
2.4 Estudio sobre la fonética española	16
2.4.1 Introducción al español	16
2.4.2 Estudio sobre el español	17
2.5 Estudio sobre la audición	18
2.5.1 Sistema auditivo	18
2.5.2 Espectro audible	19
2.6 Problemas auditivos	23
2.6.1 Clasificación pérdidas de audición	24

2.6.2 Ruido	25
2.7 Audiometría verbal	26
2.8 Medidas de calidad	27
2.8.1 PESQ	27
2.8.2 SNR	28
2.8.3 ALcons	28
3. DISEÑO	30
3.1 Elección de las herramientas	30
3.1.1 Matlab	30
3.1.2 Speech Notes	30
3.1.3 Listas	31
3.1.4 Simulador	32
4. EXPERIMENTO	33
4.1 Código	33
5. ESTUDIO	41
5.1 Estudio analítico	41
5.2 Estudio experimental	45
6. RESULTADOS	48
6.1 Análisis de los resultados	48
6.2 Conclusiones	50
6.3 Líneas futuras	51
7. MARCO REGULADOR	53
8. ASPECTO SOCIOECONÓMICO Y PLANIFICACIÓN	57
8.1 Entorno socioeconómico	57
8.1.1 Entorno social	57
8.1.2 Entorno económico	57

8.2 Presupuesto	58
8.3 Planificación	60
9. Bibliografía	64

Abstract

1. Introduction and objectives

Nowadays, hearing loss affects more people worldwide, whether due to genetic causes, age or noise. This is causing the population to have hearing problems at an earlier age than in old times when the seniors used to be those who suffered from this disability. Suffering severe hearing problems progressively hinders communication, as, at some point, the intelligibility of words gets lost.

The main objective of this project is the study of the intelligibility of simulated audios with different hearing losses. It would be very useful to be able to accurately diagnose the hearing losses in order to make the audio records so that they are adapted to each person. A hearing loss simulator implemented in Matlab is used to obtain an original audio, 3 audio files each with a specific hearing impairment. And then, to study the intelligibility of each of them by means of two independent studies, an analytical study and an experimental study, and afterwards, conclusions have been extracted, studying the results obtained.

The analytical study is performed by analyzing different quality measures in the audios simulated with hearing loss while the experimental study uses a speech recognition software for the study of the intelligibility of words.

2. Development of the project

In the first place, we clarified the goal and established work lines to achieve the proposed objective. To this end, we had to make an extensive study on different topics.

2.1. Theory

Sound is a vibration that propagates as an audible wave through space. This is perceived by the human ear when the oscillations that occur in the air pressure become mechanical waves through the human auditory system.

The human ear has three main parts, the outer, middle and inner ear. The outer and middle parts are responsible for carrying the sound wave, while the main

function of the inner ear is to process the sound to the brain through different sensory organs.

In the inner ear, the frequencies are identified by the basilar membrane, which functions as a “mechanical filter”. It separates the frequency components of complex sounds and increases the resolution of the frequency in order to be able to differentiate the sounds.

The wide spectrum sounds are not perceived by humans. We are only able to listen to frequencies in the range of 20 Hz to 20,000 Hz. Below 20 Hz, we find the infrasounds and above 20,000 we find ultrasounds, used for example in medical devices such as the ultrasound.

Another aspect to consider is the spectrum of the human voice, Mostly, frequencies in the range of 50 Hz and upwards are generated in human speech. This range, concentrates between 300 Hz and 3 KHz. With functional testing and considering the human ear spectrum, the frequency range of 300 Hz to 3400 Hz has been found the most important for speech intelligibility and speech recognition. This range of frequency is therefore the telephone bandwidth.

Each vowel can be characterized by the sum of sound waves whose frequencies fall on bands of frequencies, called formants. Formants are the fundamental frequencies of the spectral vocal tract filter. In other words it is a concentration of energy at a given frequency. The human ear distinguishes the phonemes, since it is able to differentiate the formants,

The absolute threshold of hearing (ATH) or auditory threshold is the minimum sound level of a pure tone that an average human ear with normal hearing can hear with no other sound present. It is related to the sound that can just be heard by the organism. This depends on the frequency, as low frequencies require a greater sound pressure so that the sound can be perceived.

There are different types of hearing loss: conductive, sensorineural or mixed. Conductive hearing loss may be due to problems in both, the outer and middle ear, where the main problem is that the sound cannot be properly transported to the

inner ear. Furthermore, sensorineural hearing loss is caused by problems in the inner ear or neuronal pathways that connect the inner ear to the brain.

3. Experiment

The experiment executed in this project simulates a list of phonetically balanced audios with different severe hearing loss. In this study three disabilities have been established, the first attenuates values above 1000 Hz, the second attenuates above 500 Hz and the latter attenuates above 312 Hz. These frequencies have been chosen because they are representative in the hearing range.

The results have been validated through different measures of audio quality. Thanks to the results obtained, it was possible to establish the intelligibility of the simulated audios used in this study

In speech communication, intelligibility is a measure of how comprehensible speech is in given conditions. Intelligibility is affected by the level and quality of the speech signal, type and level of background noise, reverberation (some reflections but not too many), and, for speech over communication devices, the properties of the communication system. The concept of speech intelligibility is relevant to several fields, including phonetics, human factors, acoustical engineering, and audiometry.

The first measure to take into account is PESQ, Perceptual Evaluation of Speech Quality, that is a family of standards that measure the speech quality. PESQ 862 provides an estimation of the degradation of speech quality from end to end, serving as an objective method for predicting subjective quality.

The second measure we use is the signal-to-noise ratio (SNR) which compares input and output energy of the signals. The input signal is the untreated signal with a hearing loss and the output signals we use are the signals that we simulated by adding hearing losses.

STI, Speech Index Transmission, is a quality measure of speech transmission. STI measures some physical characteristics of the channel and expresses the capability of the channel to transmit the characteristics of a voice signal. STI is an objective

predictor of quality. Related to STI, we can find Alcons, this is the last measure we use to analyze the results of this project. This measure indicates the percentage of successes in the understanding of human speech. Alcons is related to STI.

To analyze the results we will also make use of phonetics. Phonology deals with the invariant elements of the sound waves that we perceive and produce. There are different kinds of consonants, depending on where the articulation occurs, whether they are voiced or unvoiced, etc.

The human voice is produced when the vocal cords vibrate, this is only valid for sound signals, since there are deaf phonemes in which the vocal cords do not vibrate, resulting in a wave that is the combination of several frequencies, each with their corresponding harmonics. The vibration of the tonal strings determines the tone and intensity of speech.

The project analysis is based on simulated vocal audiometry. A vocal audiometry is a test that uses a series of words to determine the correct functioning of the patient auditory system. Vocal audiometry is complementary to tonal audiometry. This allows us to establish the absolute threshold of the person's hearing.

Matlab is a mathematical simulation software, which has been used to simulate hearing loss for this project. It is developed to perform complex numerical calculations. The program integrates different functionalities like matrix calculation, signal processing and graphic visualization among others. It also allows the use of Toolboxes, allowing different functionalities for specific areas.

Another application we have used is SpeechNotes, an online speech recognition application that offers a 90% success. To verify that the application's interpretation is correct, we performed a parallel search on Google.

The audios treated are lists prepared by the philology department of the Autonomous University of Barcelona. The lists of 10 words each are made exclusively to predict intelligibility.

The experiment consists of simulating hearing losses in the audios described above. For this we used a hearing loss simulator owned by Alejandro Uriz, a

researcher at the Mar del Plata University. The simulator is based on an implementation of the MPEG-1 Audio Layer 3 model, which is an audio compression standard based on the human auditory model and works by discarding the information that it fails to perceive. Thus, the auditory basal model can be switched to any specific hearing disability version, so we can obtain more reliable results with the tests.

In this study three disabilities have been established, the first attenuates values above 1000 Hz, the second of which attenuates above 500 Hz and the latter attenuates above 312 Hz. These frequencies have been chosen because they are representative in the spectrum of the human voice.

In the first place we carried out the analytical experiment, which consists of that by means of numerical calculations of the measures described previously in Matlab we have been able to take data for its later analysis. We calculated for each disability the PESQ measurement, the signal-to-noise ratio with the original audio and *Percentag Articulation Loss of Consonants* (ALcons). The results obtained are consistent. As the loss of simulated hearing loss in audios worsens, the intelligibility of the word decreases.

Then we analyzed the signal-to-noise ratio, comparing the original signal with the 3 simulations and it is observed that with the case 1 disability, a higher signal-to-noise ratio is obtained, which implies that the noise between both signals is smaller than with the disability case 3, with which a much lower signal-to-noise ratio was obtained.

Finally, we studied the percentage of joint loss of consonants, the results obtained here are close to the scale shown in the following table.

Unexceptable	Poor	Fair	Good	Excellent
100 – 33%	33 – 15%	15 – 7%	7 – 3%	3 – 0%

The experimental study consisted of recognition by an online application of speech recognition. For this, a good quality speaker has been used in which the audios have been reproduced. All of them at the same distance and at the same volume and in the same environmental conditions.

The results have been satisfactory since they resemble the results obtained theoretically in the analytical experiment. It can be seen that as the hearing loss increases, voice recognition worsens. For example, in case 1 disability it is obtained that 174 words out of 200 have been heard correctly, whereas in case 3 disability only 53 words out of 200 have been correctly perceived.

On the other hand, the human being identifies a word suffices with the consonants, as they contain more information than the vowels even though they have less energy. Consonants are distributed throughout the spectrum of the human voice, so depending on the hearing disability a person may have more or less the compression of consonants.

In order to continue with the study presented in this project you can establish different work lines to follow that we will describe down below.

The first work line to follow is to verify the results of this project in an experimental way. To do so, it would be necessary to have a certain number of voluntary evaluators who suffer the same or similar hearing losses and to conduct a study with the same audios in order to be able to compare both results in the future.

Another work line to follow is to perform a study on the intelligibility of completed sentences. The experiment would simulate hearing loss in audios and measure intelligibility, analytically and experimentally, as has been done in this project.

The study of intelligibility is complex, although it is possible to simulate possible causes of hearing loss. Every person with a hearing impairment has a different threshold for hearing. Environmental conditions must also be taken into account, since noise from other transmitters can difficult communications.

The study of this subject is necessary to help people with disabilities, for example, conducting experiments and studies on how they may be affected in different situations, and helping improve and customize hearing aids for a patient's specific disability.

4. Planification

The development of this project has involved several phases of study and learning for its proper development. From the beginning it was very clear the subject to follow and thanks to its development it has managed to expand the knowledge it had and establish new ones.

First, a general study was conducted to decide the main line of work to follow. To this end, extensive research has been done and different courses related to the subject have been carried out in order to expand knowledge.

As for the simulation, an exhaustive investigation was followed to find what we needed to make this project a reality. With the permission of the author, we were able to have the Matlab code and proceeded to adapt it to the needs of the studio and our audios.

5. Conclusion

Thanks to this experiment, a person without any hearing impairment can experience what it would be like to suffer from hearing loss. People without disabilities often do not understand how frustrating it can be for people with disabilities, since in the most common situations it can be very difficult to follow the conversation thread in addition if the disability adds background noise. This is why many people with severe hearing impairments are forced to focus too much on a conversation or even to read the lips of the person they are listening to in order to continue the conversation.

As a final conclusion, the realization of this project has been an enriching experience for the author of this project as the choice of the project's theme comes from personal experience and has served to consolidate knowledge acquired throughout the study of Communications Systems Engineering and acquire new knowledge and more interest in the subject of acoustics in general.

Capítulo 1

1 Introducción y objetivos

Presentamos aquí las motivaciones que han conducido a la redacción del presente proyecto de investigación, comentamos los objetivos que se pretenden conseguir, y se detalla la estructura a seguir para el completo y correcto desarrollo del estudio.

1.1 Motivaciones

Es importante diagnosticar, de forma precisa, las pérdidas de audición para poder conformar los registros de audio necesarios, de manera que recojan las características de la discapacidad de cada persona. Para el estudio y análisis de las discapacidades auditivas, se encuentran grandes dificultades, debido a la gran variabilidad y subjetividad de los distintos afectados, y por el propio desarrollo de estas discapacidades, que pueden originarse tanto desde problemas físicos como neuronales, que complican la medición y categorización de las pérdidas auditivas, y que, creemos, puede mejorarse profundizando en la línea de investigación que, en este proyecto, desarrollamos.

El porqué de este proyecto es consecuencia directa de la propia experiencia, donde una leve discapacidad auditiva ha marcado el desarrollo personal. Las personas con pérdidas de audición, ya sean de un oído o de ambos, a veces, se encuentran en situaciones incómodas al no poder percibir correctamente algunos tonos de frecuencia, resultando que, las situaciones más comunes, como una conversación en un bar, sean frustrantes ante la incapacidad de percibir al 100% las palabras de una conversación.

Cada día más jóvenes sufren de pérdidas de audición debidas a una variada casuística. El oído humano se deteriora con los años, pero, en el caso de la población más joven, el hábito, muy extendido, de escuchar la música muy alta, especialmente con el uso de auriculares, está ocasionando un incremento notable en las pérdidas de audición leves.

1.2 Planteamiento del problema

El objetivo principal que se afronta en este proyecto es por tanto, la medición de la inteligibilidad de las palabras en audios simulados asemejando distintas pérdidas de audición. Habitualmente, para el proceso de experimentación de las pérdidas de audición, es necesario recurrir a evaluadores humanos, que en número suficiente, y con pérdidas auditivas similares, nos permitan concluir un estudio completo y fiable.

En nuestro estudio, se persigue desarrollar una metodología que, a la postre, permita mejorar la diagnosis de las distintas discapacidades auditivas, y su repercusión en la inteligibilidad de las palabras, automatizando su simulación y su medida. La idea principal es que, una vez conocido el tipo de errores de percepción de un sujeto, sea posible simular, a través de un modelo, esa pérdida, refinándola iterativamente hasta que el reconocedor automático cometa los mismos errores.

Además, la simulación de las pérdidas de audición, nos permite una mejor identificación, y un mejor estudio, de los distintos tipos de distorsiones y ruidos ambientales que también afectan a la inteligibilidad.

Entendemos, y es por tanto clave en este trabajo, que, mediante un simulador de pérdidas de audición, podemos generar suficientes modelos y datos, y sobre distintas discapacidades auditivas y distintos ambientes, que, completados y comparados con la realidad de cada individuo, nos permitan extraer conclusiones útiles sobre el problema planteado y a futuro desarrollar equipos que merced a estos modelos permitan la mejor y rápida diagnosis.

La solución que proporcionamos al problema, es la utilización de un software capaz de simular las tres **discapacidades auditivas progresivas** que fijamos en el proyecto, ligera, leve y moderada. Tratando los audios originales, obtenemos los ficheros de audio simulados, desarrollamos dos experimentos, uno analítico, en el que realizaremos una serie de medidas de calidad sobre los audios, y otro experimental, en el que analizamos cómo de bien es capaz de entender, una aplicación de reconocimiento de voz, las palabras de los audios simulados con las distintas pérdidas de audición.

El desarrollo de este estudio traza un camino válido, en el sentido que hoy en día establecen las corrientes de Big Data, obtener tras múltiples simulaciones, grandes bases de datos que permitan inferir resultados plausibles a la hora de diagnosticar las distintas discapacidades, sus grados, orígenes... Y todo ello con unos ligeros inconvenientes únicamente en función de la propia mejora de las herramientas y simuladores que utilicemos. Si bien al final deberá existir una validación humana que corrobore los resultados obtenidos de las referidas bases.

1.3 Objetivos

Para ayudar a evaluar esta problemática, cada día más común, hemos planteado un proyecto cuyo objetivo principal es *la simulación de pérdidas de audición sobre una serie de audios previamente preparados para una prueba de inteligibilidad*.

Los ficheros de audio serán tratados, mediante un simulador en Matlab, de manera que emulen tres grados de discapacidad auditiva, ligera, leve y moderada ayudándonos, para ello, de software y herramientas elegidas a propósito, y que presentaremos más adelante.

Resolveremos dos enfoques, uno analítico, donde cuantificaremos la inteligibilidad de los audios simulados, con la ayuda de distintas medidas de calidad, y otro experimental, donde se valorará la inteligibilidad de esos mismos registros de audio con la ayuda de una aplicación de reconocimiento de voz.

Finalmente, discutiremos los resultados obtenidos para poder establecer unas conclusiones objetivas, y marcar unas líneas futuras a seguir.

Para el planteamiento y desarrollo, del proyecto, ha sido preciso:

1. Analizar las características principales de la audición.
2. Comprender las distintas las pérdidas de audición y sus características.
3. Realizar un estudio sobre simuladores en Matlab.
 - 3.1. Investigar sobre el tratamiento de audios en Matlab.
4. Implantar un sistema para evaluar la inteligibilidad.

- 4.1. Establecer estudio analítico mediante el análisis de medidas de calidad previamente elegidas.
- 4.2. Realizar estudio experimental con ayuda de reconocedor de voz automático.
- 5. Elaborar un análisis de los resultados obtenidos.
 - 5.1. Deducir conclusiones válidas sobre los resultados

1.4 Estructura de proyecto

Nuestro proyecto está desarrollado según los siguientes capítulos:

- *Capítulo 2:* Analizamos el Estado del Arte, e introducimos los conceptos principales necesarios para la correcta realización del estudio. En primer lugar, realizamos una introducción a la anatomía del oído humano, y a las características principales de la audición. Estudiaremos los problemas auditivos más frecuentes hoy en día, y haremos hincapié en las discapacidades que vamos a simular. Analizaremos, también, la voz humana y explicaremos, de forma resumida, las características principales de la fonética española. Por último, definiremos en qué consiste una audiometría verbal, e introduciremos las medidas de calidad que vamos a utilizar en el proyecto de manera teórica.
- *Capítulo 3:* Explica el porqué de la elección de las herramientas que hemos utilizado en el proyecto.
- *Capítulo 4:* Explica el código utilizado para la simulación y el experimento analítico.
- *Capítulo 5:* Desarrolla los estudios, analítico y experimental, conforme al objetivo marcado.
- *Capítulo 6:* Analiza los resultados obtenidos en ambos estudios, analítico y experimental, se extraen conclusiones de ambos y se comenta la validez del estudio. Por último, se plantean unas líneas a seguir en el futuro como consecuencia de este proyecto.

- *Capítulo 7:* Recoge el marco regulador que afecta a las discapacidades auditivas y problemáticas similares.
- *Capítulo 8:* Fija el entorno socioeconómico en que se desarrolla el proyecto y valora el coste de las actuaciones que han sido necesarias para su redacción. Por otra parte, describe la planificación que hemos establecido para el correcto desarrollo del mismo.
- *Capítulo 9:* Se presenta la bibliografía consultada en la preparación y redacción de este proyecto.

Capítulo 2

2 Conceptos preliminares

En primer lugar, se realizará el análisis del Estado del Arte, donde presentaremos diferentes estudios llevados a cabo sobre la inteligibilidad.

Introduciremos unos conceptos preliminares, estudiando el sonido, la voz humana y, brevemente, la fonética española, que nos servirán para establecer unos conceptos, básicos y necesarios, para el entendimiento del proyecto que nos ocupa.

Además expondremos un estudio de la anatomía del sistema auditivo humano, del espectro de frecuencias que puede recibir, y, por tanto, sus limitaciones.

Continuaremos, describiendo los problemas de audición más comunes que existen en la actualidad, y las restricciones que se presentan al intentar establecer comunicación oral con otros seres humanos, o en el reconocimiento del audio de otros.

Y, finalmente, expondremos un resumen de las medidas de calidad que establecemos para cuantificar la inteligibilidad, y la prueba que determina la inteligibilidad definitiva.

2.1 Análisis del Estado del Arte

En el análisis del Estado del Arte, presentamos dos estudios que nos parecen esclarecedores, y sumamente útiles.

El primer estudio (Bradley & Sato, 1986) data de 1986, y fue desarrollado por John Bradley. Se realizaron distintas pruebas de inteligibilidad del habla, en diferentes clases, de diferentes colegios, y con diferentes edades, (seis, ocho y once años), y aunque los resultados diferían de los estudios previos realizados en laboratorio, se consideraron adecuados al asemejarse a aquellos cuando las condiciones eran más reales.

Los resultados indican que la relación señal a ruido mayor a 15 dB no es adecuada para los niños más pequeños. Combinando los resultados de las pruebas de simulación de pérdidas de audición y medidas de inteligibilidad del habla, con las mediciones de los niveles de habla y ruido presentes en situaciones reales, se pudo establecer una estimación del porcentaje de alumnos que experimentaron condiciones acústicas casi ideales. Los resultados sirvieron para poder estimar criterios acústicos para mejorar las aulas de enseñanza de primaria.

Otro estudio a tener en cuenta es el desarrollado por Joanne Rajadurai, profesora en la universidad de Heriot-Watt, en Malasia (Rajadurai, 2007). En este estudio se examinan las cuestiones que plantean las investigaciones previas sobre la inteligibilidad, cuando una persona no se comunica en su idioma materno, en este caso el inglés. La investigación se llevó a cabo teniendo en cuenta el contexto de interacción de los participantes involucrados, y haciendo énfasis en la pronunciación de éstos. Los resultados de esta investigación han servido para avanzar la idea de que, la fonología del habla inteligible, no puede fijarse en un conjunto estrecho de características asociadas a una variedad particular del habla, sino que debe ser más flexible y ampliar las variantes que constituyan una pronunciación inteligible.

2.2 El sonido

El sonido es una onda acústica, es decir una perturbación, que se propaga por el espacio. Es recibido por el oído humano cuando las oscilaciones de la presión del aire a través del oído, son percibidas, finalmente, por el cerebro. (Crocker, 1997)

La propagación del sonido involucra el transporte de energía, y puede transmitirse a través de cualquier medio elástico bien sea sólido, líquido o gaseoso. Podemos diferenciar entre *onda transversal*, cuando las vibraciones propias del sonido son perpendiculares a la dirección de propagación, y *onda longitudinal* cuando la vibración y la dirección de propagación son paralelas. Las ondas sonoras son ondas longitudinales, aunque podemos encontrar ondas transversales en instrumentos musicales como en la vibración de una cuerda.

Para el estudio de los sonidos referidos al habla humana, tenemos la *fonética acústica* que busca explicar cómo se generan, cómo se perciben y cómo se pueden describir estos sonidos.

2.3 La voz humana

La voz humana se forma cuando las cuerdas vocales vibran produciendo señales sonoras, puesto que existen muchos fonemas sordos en los que las cuerdas vocales no vibran. En ese momento, se genera una onda sonora, combinación de varias frecuencias con sus correspondientes armónicos. (Bernal, Bobadilla, & Gómez, 2000)

Tras una serie de variaciones, en un mismo rango o frecuencia de emisión, obtenemos, como resultado, una frecuencia fundamental y sus armónicos. La vibración de las cuerdas vocales determina el tono y la intensidad del habla.

Debemos destacar los picos que se encuentran en el espectro sonoro de las vocales, y que se conocen como *formantes*. Un formante es un pico de intensidad en el espectro del sonido, es una concentración de energía a una determina frecuencia. En la figura 1 podemos ver una representación de los formantes para las vocales en español.

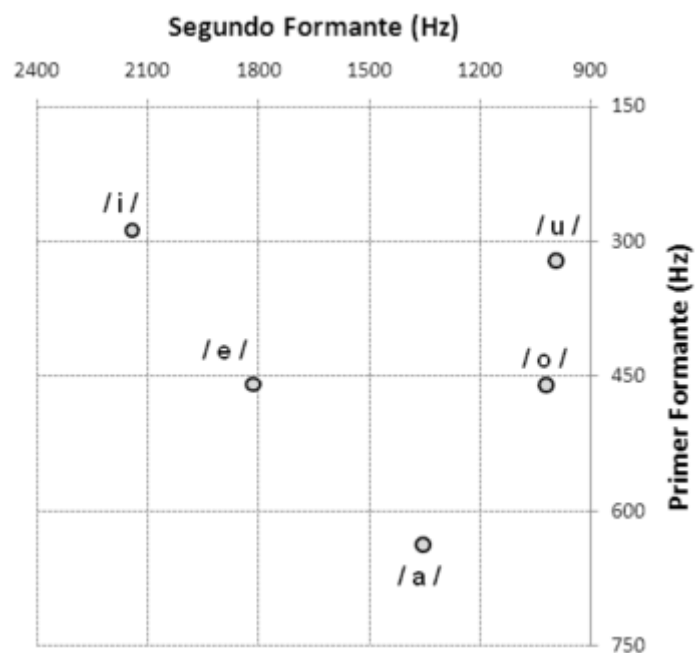


Figura 1 - Mediciones de los primeros formantes de las vocales en español (Bradlow, 1994)]

Debido a que la información espectral de la señal de voz no se distribuye de forma homogénea en frecuencias, para el procesamiento de la señal, el espectro se divide en diferentes bandas de frecuencias y disminuir así las discontinuidades entre bandas conjuntas. (Fletcher & Munson, 1933)

Es importante señalar *la banda crítica* que representa un rango de frecuencias, dentro de las cuales, un sonido bloquea o enmascara la percepción de otro sonido. *La banda crítica* describe el ancho de banda de frecuencia del “filtro auditivo” creado por la cóclea.

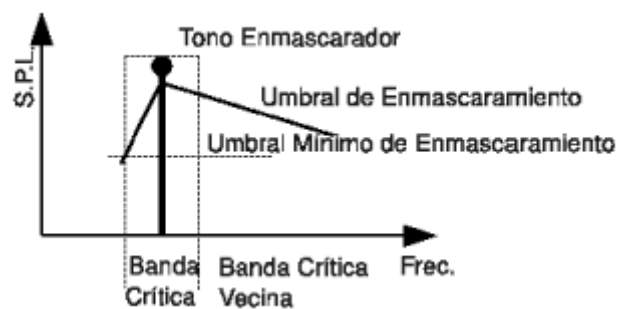


Figura 2 - Representación bandas críticas (A. Uriz)

En la figura 2, vemos una representación del tono enmascarador explicado en el apartado de ruido, y las bandas críticas descritas anteriormente.

2.4 Estudio sobre la fonética

Existen menos estudios sobre la inteligibilidad de las palabras en castellano pues, la mayoría de estudios realizados son en inglés, Por ello, se considera necesaria una breve introducción al idioma español, y a sus características fonéticas principales.

2.4.1 Introducción al español

El idioma español o castellano es una lengua romance procedente del latín hablado. Tiene su origen en la actual Castilla y León, y es una continuación moderna del latín hablado desde el siglo III, y que, tras el desmembramiento del imperio romano, fue divergiendo de las otras variantes romances como el francés, el italiano...

El español es la segunda lengua hablada en el mundo, tras el chino mandarín. Se habla en España, Hispanoamérica, el Sahara Occidental, Guinea Ecuatorial y en algunas partes de Filipinas, siendo por ello, además, uno de los seis idiomas oficiales de la ONU.

2.4.2 Estudio sobre el español

La fonología es la ciencia que se ocupa del estudio de los elementos invariantes, de las unidades lingüísticas codificadas en las ondas sonoras que producimos, y que percibimos los seres humanos.

Los fonemas, que son las unidades de las que se ocupa la fonología, contienen un carácter distintivo y producen diferencias de significado.

La fonética es la variación articulatoria y acústica de los sonidos del habla, y el modo en los que éstos son percibidos, es decir, trata sobre el estudio de la naturaleza acústica y fisiológica de los sonidos.

LAS CONSONANTES CASTELLANAS				bilabial	labiodental	interdental	postdental	alveolar	prepalatal	mediopalatal	postpalatal	velar	uvular	
obstruyente	+ cierre	oclusiva		sorda	p		t̪	t				k̟	k	
				sonora	b			d			g̟	g		
		africada		sorda						ç̟				
				sonora							j̟			
	- cierre	aproximante		sorda	β̞		ð̞		ɹ̞			ç̞	ɣ̞	
				sonora	β		ð		ɹ		y	ɣ̟	ɣ	
		fricativa	hendidura alargada	sorda	(f)	f	θ					x̟	x	X
				sonora			ð̞							
			hendidura en forma de canal	sorda				s̟	s					
				sonora				s̟	s̟	ʒ̟				
sonorante	nasal		sonora	m	m̟	ɱ	ɱ̟	n	ɲ̟	ɲ		ŋ̟		
	líquida	lateral	sonora				l̟	l̟	l		λ̟			
		vibrante sonora	simple					r̟						
			múltiple						̄r̟					

Tabla 1 - Clasificación de las consonantes del español

En la tabla 1 (Canellada & Madsen, 1987) vemos una clasificación de las consonantes del español, según sea el punto de articulación, según dónde se produce la articulación del fonema, según el modo de articulación, dependiendo de la posición que adopten los órganos durante el habla, o por la vibración de las cuerdas vocales, que pueden ser sonoras o sordas.

Y, mención especial para los formantes, que representan picos de intensidad en la envolvente espectral del sonido, permitiendo diferenciar los sonidos propios del habla humana, sobre todo las vocales, y otros sonidos sonoros, que podemos producir, cuando no existe una turbulencia en el flujo de aire al atravesar el tracto vocal, como ocurre en /m/ o /l/.

2.5 Estudio sobre la audición

2.5.1 Sistema auditivo

El sistema auditivo humano constituye una parte esencial en la comunicación humana. Se encarga de convertir las ondas sonoras en señales eléctricas que se transmiten, a través del nervio acústico, hasta el cerebro donde el sonido es interpretado.

El oído humano consta de tres partes principales: oído externo, oído medio y oído interno. Las partes externa y media transportan las ondas sonoras hacia el oído interno, en el que están los órganos sensoriales del sistema auditivo interno. Como podemos apreciar la estructura en la figura 3.

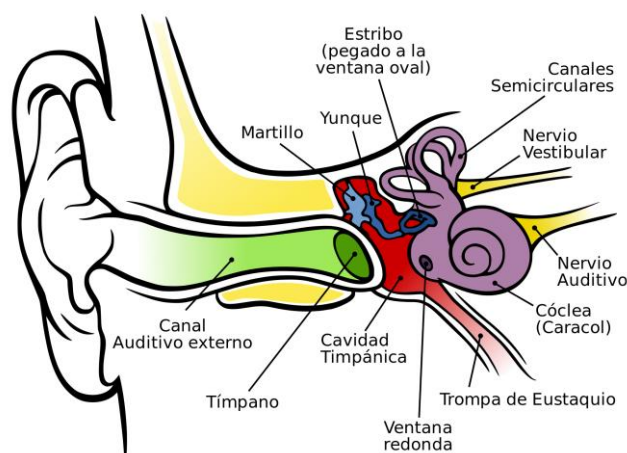


Figura 3 - Representación del sistema auditivo humano. (Padial, 2017)

El oído externo, está compuesto, primero, por el pabellón auditivo, conocido comúnmente como oreja, que aumenta la frecuencia y localiza la fuente sonora, y en segundo lugar, por el conducto auditivo externo, que se encarga de transportar la onda recogida hacia la membrana timpánica o tímpano. Su principal función es captar el sonido y transmitirlo al oído medio.

El oído medio es la cavidad pequeña que se encuentra en el interior de la porción petrosa del hueso temporal, y está limitada por el tímpano y la base de la cóclea. En su interior, podemos encontrar tres huesos pequeños, denominados martillo, yunque y estribo, y el tubo auditivo o trompa de Eustaquio. Su misión es convertir las ondas sonoras en transmisión mecánica hasta el oído interno.

El oído interno, también denominado laberinto o cóclea, se encuentra dentro del hueso temporal, y contiene los órganos sensoriales de la audición y el equilibrio. Dentro de la cóclea, podemos señalar la membrana basilar, que modifica su rigidez en función de la frecuencia de la onda sonora. Esta membrana es un auténtico banco de filtros mecánicos que separan las componentes frecuenciales de los sonidos complejos. La función principal de estos filtros es aumentar la resolución de frecuencia de la cóclea para que, así, podamos diferenciar los distintos sonidos que finalmente interpretará el cerebro. (Borja, 2011)

2.5.2 Espectro audible

El oído del ser humano posee limitaciones para percibir sonidos, pues solo puede escuchar los sonidos emitidos en un rango de frecuencias muy específico; su límite inferior se encuentra en 20 Hz, y su límite superior lo fija la frecuencia de 20.000 Hz, dentro de un espectro de frecuencias que mucho más amplio, y caracterizado por la distribución de amplitudes para cada frecuencia del fenómeno ondulatorio, ya sea sonoro, luminoso o electromagnético. Nuestro estudio se centra en el fenómeno sonoro.

Una fuente de ondas sonoras suele ser una superposición de frecuencias diferentes. Cada frecuencia estimula una parte distinta de nuestra cóclea de manera que, el oído humano, al escuchar una onda sonora con una única frecuencia dominante percibe una nota, pero si tal dominante no existe, percibirá

frecuencias dentro del todo rango audible. En la figura 4, vemos una representación de las diferentes partes estimuladas en la cóclea dependiendo de la frecuencia.

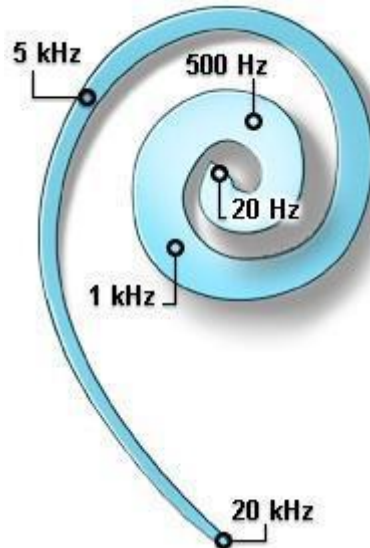


Figura 4 – Representación frecuencias en la cóclea. (Pujol, 2016)

En el análisis de la voz humana debemos destacar un rango de frecuencias que se sitúa desde 300 Hz hasta 3400 Hz aproximadamente. Esta banda es la que contiene la mayor parte de energía espectral de la voz humana. Por ejemplo, cada vocal se caracteriza por la suma de ondas sonoras cuyas frecuencias se encuentran dentro de bandas de frecuencias denominadas formantes. El oído humano es capaz de distinguir las vocales, como consecuencia de poder discriminar dichos formantes, es decir, comprende parte del espectro de frecuencias presentes en una onda sonora que produce la articulación de dicha vocal. Esta banda al ser significativa en la voz humana, es la utilizada en un canal telefónico.

Existen otros rangos de frecuencias, aunque no son perceptibles para el oído humano. Tenemos *los infrasonidos*, sonidos de una frecuencia inferior a 15 Hz, inaudibles por el oído humano, aunque podemos percibir sus vibraciones en los tejidos blandos del cuerpo humano como el abdomen. Y los *ultrasonidos*, sonidos con frecuencias superiores a 20 KHz, y que pueden ser escuchados por algunos animales como los perros. Para los ultrasonidos no está definido un límite superior.

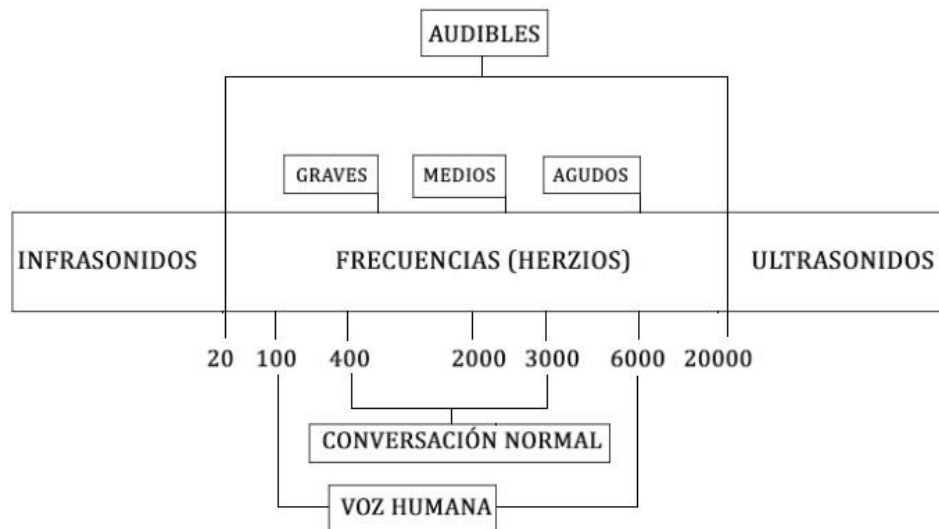


Figura 5 - - Representación del espectro de frecuencias de los sonidos

En la figura 5 se recoge una representación del espectro de frecuencias de los sonidos, señalando claramente que, para que un sonido sea audible para las personas, debe estar contenido dentro del rango de 20 Hz a 20 KHz.

Un punto importante a tener en cuenta en el estudio de la audición humana, es el *umbral de audibilidad o umbral de audición*, que marca la intensidad mínima de sonido capaz de impresionar al oído humano; en otras palabras, es el nivel mínimo de un sonido para que logre ser percibido. Habitualmente se expresa en dB referidos a una presión de 20 μ Pa, y depende de la frecuencia de manera que, a frecuencias bajas, es imprescindible una presión sonora mayor para producir una sensación audible similar a la que se produciría a 3000 Hz donde el umbral es menor.

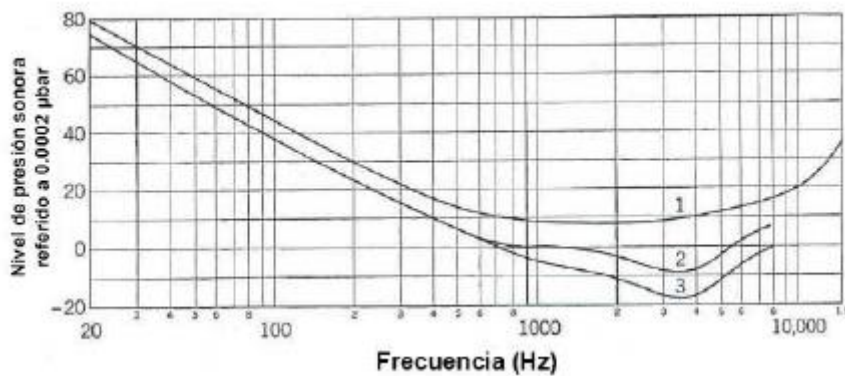


Figura 6 -- Umbral de audibilidad monoaural y binaural

En la figura 6 (Perez & Zamanillo, 2003) podemos observar que, el umbral de audibilidad, depende de la frecuencia pudiendo establecerse tres curvas. A saber:

- La curva 1 que representa el umbral de audibilidad monoaural, es decir, para un único oído, y la fuente sonora junto a él. En este caso, la presión sonora está medida a la entrada del canal auditivo.
- La curva 2 comprende el umbral binaural, y la fuente sonora está enfrente del oyente.
- La curva 3 se corresponde con un umbral binaural cuando la fuente sonora está constituida por diversas fuentes sonoras distribuidas de forma aleatoria en el plano horizontal alrededor del oyente.

El umbral de audibilidad está ligado a muchos factores, pues, para una misma persona, puede variar dependiendo del entorno y las circunstancias ambientales en las que se encuentre. Relevante, dentro de estas variables que afectan al umbral de audibilidad, es la edad de los individuos pues, generalmente, a mayor edad, menor umbral auditivo.

En la figura 7 podemos ver los límites del umbral de audición de las personas, y en la figura 8 se representa la curva estándar del mismo simulada con la herramienta Matlab.

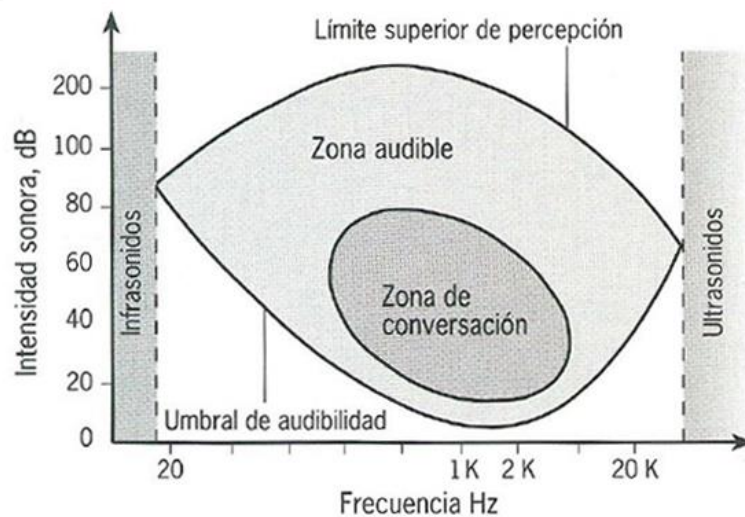


Figura 7 - Gráfico que representa el umbral de audibilidad (Borja, 2011)

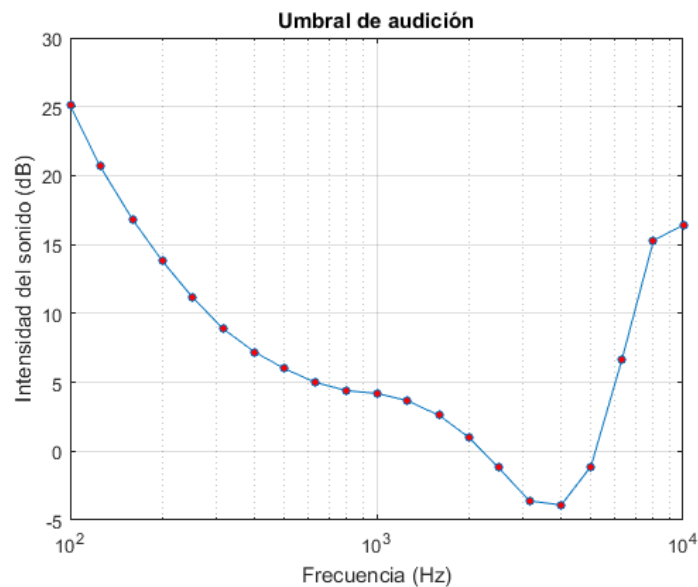


Figura 8 - - Umbral de audición estándar simulado en Matlab

2.6 Problemas auditivos

Importante dentro del problema que nos ocupa, es establecer los problemas auditivos y la índole de éstos. Podemos establecer tres tipos principales de pérdidas auditivas atendiendo a la localización de éstas, tenemos así:

- la conductiva,
- la neurosensorial, y
- la mixta.

, y podemos clasificar estas pérdidas de audición,

- según el grado de pérdida auditiva, o
- según la localización de la zona afectada

2.6.1 Clasificación de pérdidas de audición

Según la localización de la lesión:

- *Pérdida auditiva conductiva*, debida a problemas físicos, tanto en el oído externo como en el oído medio, y que no permitan transmitir adecuadamente el sonido. Puede ser temporal, y se puede corregir mediante el uso de un audífono o un implante de oído medio.
- *Pérdida auditiva neurosensorial*: Se produce cuando hay daño en el oído interno, en la cóclea, o en los conductos entre el oído interno y el cerebro. Esta pérdida suele ser permanente.
- *Pérdidas mixtas* que incluyen tanto daños conductivos, como neurosensoriales (daño en el oído externo o medio, y en el interno o en el nervio auditivo).

Según el grado de pérdida auditiva:

Los grados de pérdida auditiva se establecen, de forma general, en leve, moderada, severa y profunda, atendiendo a la severidad de la pérdida. En la figura 9, se presenta una tabla, de uno de los sistemas más comunes de clasificación, donde el intervalo en dB muestra el rango de pérdida de audición.

Grado de pérdida de audición	Escala de la pérdida de audición (dB)
Normal	-10 a 15
Ligera	16 a 25
Leve	26 a 40
Moderada	41 a 55
Moderadamente severa	56 a 70
Severa	71 a 90
Profunda	91+

Tabla 2 - Representación escala de pérdidas de audición (Clark, 1981)

Donde podemos establecer como principales los grados de:

- *Leve*: Incapacidad para escuchar sonidos inferiores a 30 dB. Si existen ruidos de fondo, dificulta la comprensión del habla.
- *Moderada*: Incapacidad para escuchar sonidos inferiores a 50 dB. Sería necesario un audífono.
- *Severa*: Incapacidad para escuchar sonidos inferiores a 80 dB. En este caso, puede que ni con un audífono sean capaces de escuchar por lo que deberán comunicarse en lengua de signos o mediante lectura labial.
- *Profunda*: Incapacidad de escuchar sonidos inferiores a 95 dB. Es necesaria la comunicación mediante lengua de signos o lectura labial.

2.6.2 Ruido

Como uno de los causantes principales de las pérdidas auditivas está el ruido. El ser humano se encuentra continuamente expuesto a sonidos ambientales que no afectan a su audición. Sin embargo, los ruidos muy altos, pueden causar graves daños ya sean temporales o permanentes. La exposición a niveles dañinos de ruido puede darse a cualquier edad, y en cualquier momento; baste comentar que el 15% de los adultos en Estados Unidos, o sea 26 millones de personas entre los 20 y 69 años, sufren una pérdida de audición debida a la exposición al ruido en sus actividades recreativas o de trabajo. (National Institute on Deafness and Other Communication Disorders (NIDCC), 2017)

El ruido afecta de diferentes maneras a la percepción del habla humana. Podemos, no obstante, distinguir tres casos principales (Miyara, 2017):

- *Alteración del espectro*: Ocurre cuando un ruido se superpone a un sonido útil y, como resultado, el espectro difiere del original, y, dado que la percepción comienza con un análisis del espectro, resulta un patrón espectral alterado que dificulta la interpretación del habla.
- *Enmascaramiento*: Se produce cuando hay dos sonidos, y uno de ellos, presenta una intensidad bastante mayor (unos 20 ó 30 dB mayor). En este caso, el de menor intensidad, no se llega a percibir. Esta limitación, que

puede resultar útil, pues permite al cerebro centrarse en un ruido obviando sonidos insignificantes, como por ejemplo una conversación lejana, se convierte en un problema cuando el sonido enmascarado se refiere a la palabra hablada; en este caso, se produce una *pérdida de inteligibilidad*. El enmascaramiento puede ser total o parcial. Si es total, se enmascararía toda la emisión vocal, es decir, cuando el ruido supera al habla humana. En el caso del enmascaramiento parcial, los sonidos débiles son los que sufren el enmascaramiento de las pequeñas diferencias que permiten distinguir las consonantes. El enmascaramiento parcial también afecta a la inteligibilidad pues los sonidos que primero se pierden, las consonantes, son los que transportan la mayor parte de la información (si se confunde la letra “s” por una “j” en la palabra casa, cambiaría totalmente el significado de la palabra).

- *Confusión de patrones temporales*: Este caso se da cuando un ruido o impacto sobrepasa a una emisión vocal, el problema surge cuando consonantes con un perfil temporal similar como la “c” y la “t” pueden ser confundidas.

2.7 Audiometría verbal

La *audiometría verbal* o *logoaudiometría*, es una prueba en la que se utilizan palabras para determinar el correcto funcionamiento del sistema auditivo. La audiometría verbal es complementaria de la audiometría tonal, la que permite conocer los umbrales absolutos que percibe cada sujeto. (UAB, 2017)

Esta prueba, que permite establecer la inteligibilidad del habla del sujeto que la realiza, se muestra en porcentaje de respuestas correctas, de manera que, el umbral de inteligibilidad, está representado por el nivel que incluya, al menos, el 50% de acierto de los elementos o palabras presentados. En la figura 9, vemos que la *curva A* se considera normal, es decir, cuando el sujeto es capaz de acertar todas las palabras. La *curva B* representa una hipoacusia o disminución de la sensibilidad auditiva, mientras que la *curva C* muestra una pérdida severa de inteligibilidad, con la aparición de distorsiones importantes para intensidades mayores a 80 dB HL.

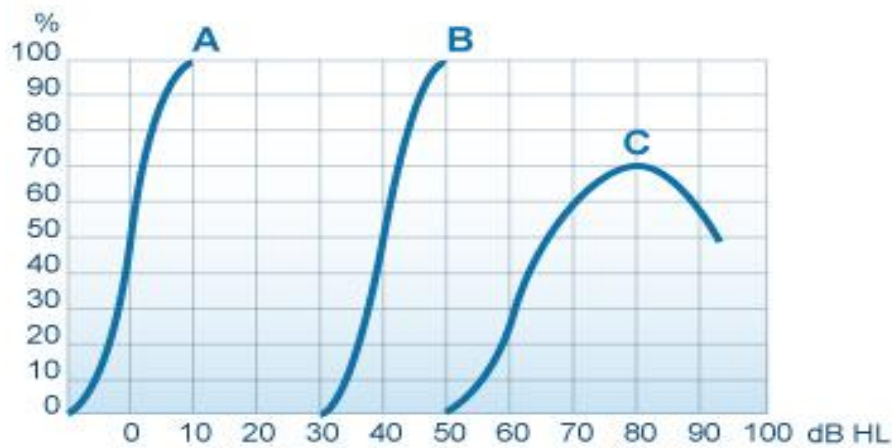


Figura 4 - Representación curvas de inteligibilidad (Chaix, 2016)

2.8 Medidas de calidad

Sabemos que la inteligibilidad de una palabra es un aspecto subjetivo, pero podemos encontrar un método analítico para obtener, cuantitativamente, unos porcentajes estadísticos de ésta.

En el proyecto hemos establecido una serie de medidas para poder evaluar la calidad del habla humana, que nos permitan realizar una cuantificación de las pérdidas de audición, y establecer unos resultados sobre éstas y, consecuentemente, la inteligibilidad. En definitiva, estas medidas, nos permitirán establecer cómo de buena es la calidad del sonido.

Utilizaremos las siguientes herramientas de medida:

2.8.1 PESQ

La medida PESQ, *Perceptual Evaluation of Speech Quality*, de extremo a extremo en redes telefónicas de banda estrecha y códecs vocales, describe un método objetivo para poder predecir la calidad subjetiva de una señal. PESQ compara una señal de entrada, con una señal degradada, de manera que, la salida de PESQ, resulta ser una predicción de la calidad percibida por los sujetos en una prueba de escucha subjetiva, y que sería atribuida a la señal degradada. (P.862, 2001)

Como desventaja de esta medida, se debe tener en cuenta que, el algoritmo PESQ, no proporciona una evaluación exhaustiva de la calidad de transmisión, sino que mide solamente los efectos del ruido y la distorsión de voz unidireccional sobre la

calidad vocal; no obstante, para el experimento que realizamos en este diseño, lo tomaremos como medida objetiva. El rango válido de esta medida es de 0 a 5, siendo 5 una calidad de transmisión excelente y 0 una calidad inaceptable.

2.8.2 SNR

La SNR, *Signal Noise to Ratio*, o relación señal a ruido, es una medida que define la relación entre la potencia de una señal con la potencia de ruido que la corrompe, es decir, compara el nivel de una señal deseada con el nivel del ruido de fondo; cuánto más alta es la relación, menos molesto es el ruido de fondo.

$$SNR = \frac{P_{señal}}{P_{ruido}}$$

Ambas potencias, de la señal y del ruido, deberán ser medidas en un mismo sistema equivalente, y en el mismo ancho de banda, para obtener un correcto resultado. Como las diversas señales pueden presentar un rango dinámico bastante amplio, las potencias se expresan, principalmente, en la escala logarítmica. La potencia de la señal original se puede expresar en dB, como recoge la ecuación siguiente, y, de forma análoga, puede expresarse para la potencia de ruido. La relación señal a ruido suele representarse en decibelios.

$$P_{señal_{dB}} = 10 \cdot \log_{10} P_{señal}$$

Una desventaja a tener en cuenta, en señales digitales, es que, la relación señal a ruido, se verá reducida si no se emplea un número suficiente de niveles de cuantificación, además, de estar limitada por el criterio de Nyquist, donde la frecuencia de muestreo debe ser el doble de la frecuencia máxima de la señal. (Loizou, 2007)

2.8.3 ALcons

El índice de articulación, ALcon (*Articulation Loss of Consonants*), permite indicar el porcentaje de aciertos en la comprensión del habla humana. Podemos diferenciar 3 tipos de índices:

- *el índice de articulación silábico*, donde el usuario debe escuchar una cierta cantidad de sílabas sin relación entre ellas.

- *el índice de articulación de palabras*, para el que, en la escucha, se utilizan palabras, normalmente de dos sílabas, y
- *el índice de articulación de frases*, que utiliza frases completas en la audición.

En todos los casos, *las emisiones son fonéticamente balanceadas*, es decir, los fonemas aparecen en la misma proporción que en el habla normal.

En nuestro estudio, nos hemos centrados en el índice de articulación de palabras, en concreto, en el porcentaje de consonantes que se pueden identificar en los audios tratados experimentalmente.

Este método evalúa la pérdida de definición de las consonantes, estando este criterio relacionado profundamente con el grado de inteligibilidad de manera que se establece que:

- 0-10% →Inteligibilidad muy buena
- 10-15% →Inteligibilidad buena
- >15% →Inteligibilidad insuficiente

Estos márgenes, no obstante, están basados en estadísticas, y, por tanto, son variables, pudiéndose dar el caso de que una persona no comprenda o diferencie palabras con una ALcons del 15%, lo que abunda en la amplia variabilidad del caso que afrontamos.

Esta medida se relaciona con STI, *Speech Transmission Index*, (Donley, 2015) que es una medida de la calidad de la transmisión del habla, y que hace uso de la función de transferencia de modulación. La STI mide algunas características físicas de un canal de transmisión, y expresa la capacidad del canal para transmitir las características de una señal específica de voz. Es un predictor objetivo, tanto de la medición, como de las características del canal de transmisión, que afectan a la inteligibilidad del habla. La relación matemática entre ambas medidas se puede expresar, aproximadamente, como:

$$ALcons = 10^{\frac{1-STI}{0.46}}$$

Capítulo 3

3 Diseño

En este capítulo explicaremos las herramientas elegidas para el desarrollo del proyecto y detallaremos la elección de cada una y su adecuación para conseguir los objetivos fijados.

3.1 Elección de las herramientas

3.1.1 MATLAB

Matlab es el software de simulación que hemos utilizado para realizar este proyecto. MATLAB, *MATrix LABoratory*, es una herramienta de software matemático que ofrece un entorno de desarrollo integrado con un lenguaje de programación propio (.m). Está orientado para el desarrollo de elevados cálculos matemáticos, y la visualización de éstos. MATLAB integra distintas funcionalidades como cálculo matricial, análisis de números, procesamientos de señales y visualización gráfica, y todo ello, en un entorno completo utilizando el mismo lenguaje. Además, integra distintas utilidades denominadas, Toolboxes, para el apoyo al usuario. Actualmente existen Toolboxes para cada área del campo de la ingeniería y de la simulación.

La elección de Matlab para este proyecto ha sido por las herramientas que ofrece en el tratamiento de audio, en particular en este estudio se ha hecho uso de la *Auditory Toolbox* (Slaney M. , 1998). Esta herramienta implementa varios modelos auditivos, siendo útil para el estudio del habla y la audición, y para ver cómo el sistema auditivo humano representa esos sonidos.

3.1.2 Speech Notes

Speech Notes es la aplicación informática de reconocimiento de voz que hemos utilizado para escuchar los audios simulados. Está basada en motores de reconocimiento de voz de alta calidad proporcionados por Google, y ofrece, entre otras posibilidades, y ésta es la que nos interesa, un bloc de notas en línea. Para asegurar la mejor interpretación, todo lo que escucha, se envía al buscador,

utilizando algoritmos paralelos, y es devuelto y enviado a Speechnotes como un flujo de posibles resultados de transcripción, con unos niveles de precisión superiores al 90%. La elección de esta aplicación frente a otras posibles, como *Dragon Naturally Speaking*, es consecuencia de su gratuidad y, no obstante, cumplir suficientemente con los requisitos marcados en este proyecto.

3.1.3 Listas

Los audios utilizados en este proyecto son propiedad de Widex Audífonos SA (1993). Se trata de un CD de audio preparado para poder realizar distintas pruebas de audición. Las listas proporcionadas en él son obra del Departamento de Filología Española de la Universidad Autónoma de Barcelona, siendo sus autores Juan María Garrido y Joaquim Llisterri (Widex Audífonos, 1993).

Estas listas están preparadas para la realización de pruebas de inteligibilidad con palabras. Constan de 200 palabras diferentes, distribuidas en 10 series de 20 palabras cada una. Las palabras de cada lista están fonéticamente equilibradas. El equilibrio fonético se evalúa comparando las frecuencias de aparición en la lista, y las frecuencias de aparición en una distribución de referencia.

Las características fonéticas de los audios se corresponden con:

- Palabras bisílabas.
- Patrón acentual paroxítono
- Estructuras silábicas CV, CVC y CCV (Consonante Vocal, Consonante Vocal Consonante, Consonante Consonante Vocal)

... y, en cuanto a sus características léxicas, son palabras seleccionadas entre las 3000 más frecuentes del diccionario de Juilland y Chang-Rodríguez. (Juilland Alphonse, 1964)

Existen otras bases de datos que podrían habernos servido para ejecutar la simulación, como Albayzin, que es una base de datos diseñada por un Consorcio de seis grupos de investigación y producida por la empresa Page Ibérica S.A. Esta base consta de 3 corpus de señales grabadas a 16 KHz y producidas por 304 locutores

de la variedad central del castellano, un corpus fonético, uno geográfico y un “Lombard”, una mezcla de los anteriores.

Nosotros hemos elegido los audios inicialmente comentados por ser más concisos, y su aplicación más directa, para el estudio de la inteligibilidad en las tres discapacidades que establecemos.

3.1.4 Simulador

El simulador, cedido por Alejandro Uriz, (A. Uriz) se implementa siguiendo el modelo de MPEG-I Audio Layer, estándar de compresión de audio basado en un modelo auditivo humano. Su funcionamiento se basa en el descarte de información que, él mismo, no es capaz de percibir. En la figura 10, podemos ver un modelo de un sistema genérico de codificación perceptual.

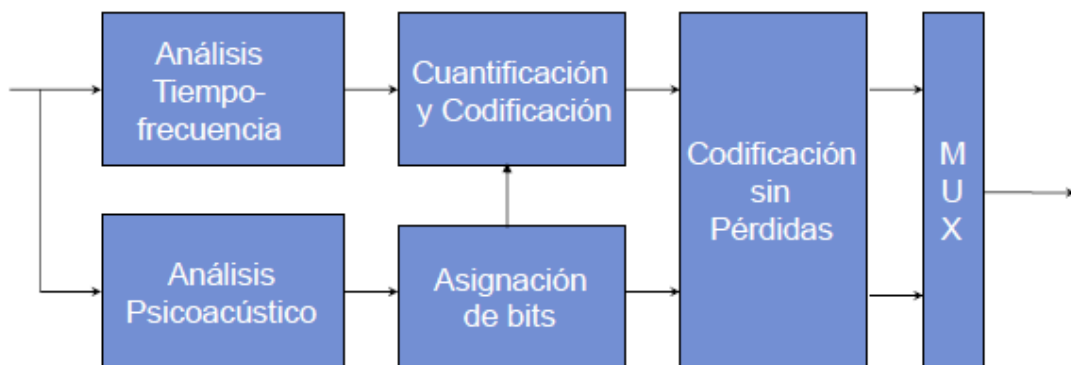


Figura 5 - Representación codificación perceptual

En nuestro caso, se modifica el modelo auditivo en el que está basado, con el objetivo de generar el sistema de audición de una persona con una discapacidad auditiva concreta, los audios generados se caracterizan así con esa pérdida de audición. En el capítulo siguiente se detallará cómo, matemáticamente, simula las distintas discapacidades que hemos establecido.

Capítulo 4

4 Experimento

4.1 Código

El simulador utilizado genera, de cada audio de entrada tratado, varios audios de salida con distintas pérdidas de audición. Nosotros nos hemos centrado en las *pérdidas de audición progresivas*, ligera, leve y moderada, que para nosotros serán caso 1, caso 2 y caso 3, respectivamente.

El software del simulador está basado en una implementación del modelo MPEG-I Audio Layer, (ISO, International Organization of Standardization, 1993) que es un estándar de comprensión de audio, basado en el modelo auditivo humano, y que funciona descartando la información que no llega a ser percibida.

El experimento completo comienza con la carga de un fichero que contiene los nombres identificativos de los audios a tratar, y, con él, trabajamos en bucle para que, cada vez que lea un nombre, trate ese audio mediante el simulador, y nos genere 4 audios de salida: el primero sin simular, y, a continuación, las simulaciones de las tres discapacidades a estudio, y cuyas características más adelante señalamos.

La función principal del simulador es *oído_solapado_or_v2_RIAI*. Esta función tiene como variables de entrada, la señal de audio, el ancho de ventana, el tiempo, el nombre del archivo de salida, y la función ATH, y devuelve, el archivo de audio tratado, con las características introducidas. La función ATH es un vector que representa el umbral absoluto de audición, y establece la curva de intensidad de presión de sonido requerida en función de la frecuencia de un ambiente sin ruido. El código de Matlab de la curva ATH está descrita en el cuadro 1, y está representada en la figura 7.

$ATH = 3.64 \cdot (f./1000)^{-0.8} - 6.5 \cdot \exp(-0.6 \cdot (f./1000 - 3.3)^2) + 0.001 \cdot (f./1000)^4;$

Cuadro 1 – Representación curva ATH

En el procedimiento que establecemos, para simular las distintas pérdidas de audición, modificamos la curva anterior para que, en nuestro caso, las personas dejen de percibir las frecuencias superiores; así, de forma gradual se pierden, primeramente, las frecuencias altas y, posteriormente, las bajas. Para poder modelar esta pérdida añadimos 50 dB, en las frecuencias superiores a 400 Hz, en la curva ATH, provocando así que, los sonidos que tengan un nivel por debajo de esa curva, no puedan ser percibidos por el oyente o se perciban de manera distorsionada. En este proyecto, según lo anterior, se han establecido para su estudio tres discapacidades auditivas. A saber:

- *caso1_discapacidad*, donde se eliminan las frecuencias superiores a 1000 Hz,
- *caso2_discapacidad*, con la eliminación de las superiores a 500 Hz y, por último,
- *caso3_discapacidad*, con la eliminación de las frecuencias superiores a 312 Hz

Estas discapacidades podemos clasificarlas como ligera, leve y moderada respectivamente, y podemos definir las con las siguientes expresiones:

```
%CASO 1  
ATH(30:length(ATH))=ATH(30:length(ATH))+50; %LPF fc=1000Hz;
```

Cuadro 4 - Representación curva ATH para discapacidad caso 1

```
%CASO 2  
ATH(20:length(ATH))=ATH(20:length(ATH))+50; %LPF fc=500Hz;
```

Cuadro 3 - Representación curva ATH para discapacidad caso 2

```
%CASO 3  
ATH(10:length(ATH))=ATH(10:length(ATH))+50 %LPF fc=312Hz;
```

Cuadro 2 - Representación curva ATH para discapacidad caso 3

Resultado de aumentar la curva ATH en 50dB, o bien realizar un corte a partir de la frecuencia que queremos fijar. Para establecer las simulaciones de las discapacidades auditivas antes señaladas, utilizamos un filtro paso bajo, *LPS Low Pass Filter*, que es un filtro caracterizado por permitir el paso de las frecuencias más bajas, y atenuar las frecuencias más altas. (Baer, 2002)

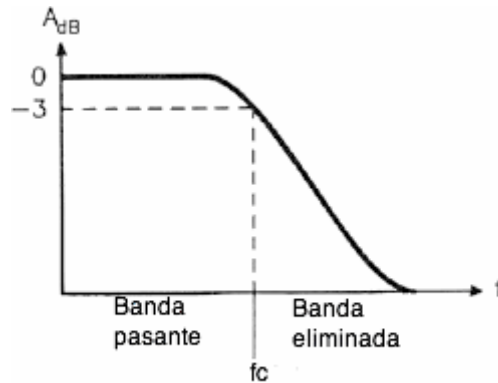


Figura 6 - Representación filtro paso bajo

En la figura 11, tenemos la representación gráfica de un filtro paso bajo, de manera que, el punto f_c , que corresponde a la frecuencia de corte, resulta ser el límite entre las frecuencias permitidas y las frecuencias atenuadas.

Y en las figuras 12,13 y 14 siguientes, podemos ver representado el umbral absoluto de audición, para cada una de las tres discapacidades simuladas, con la utilización de un filtro paso bajo como el descrito anteriormente.

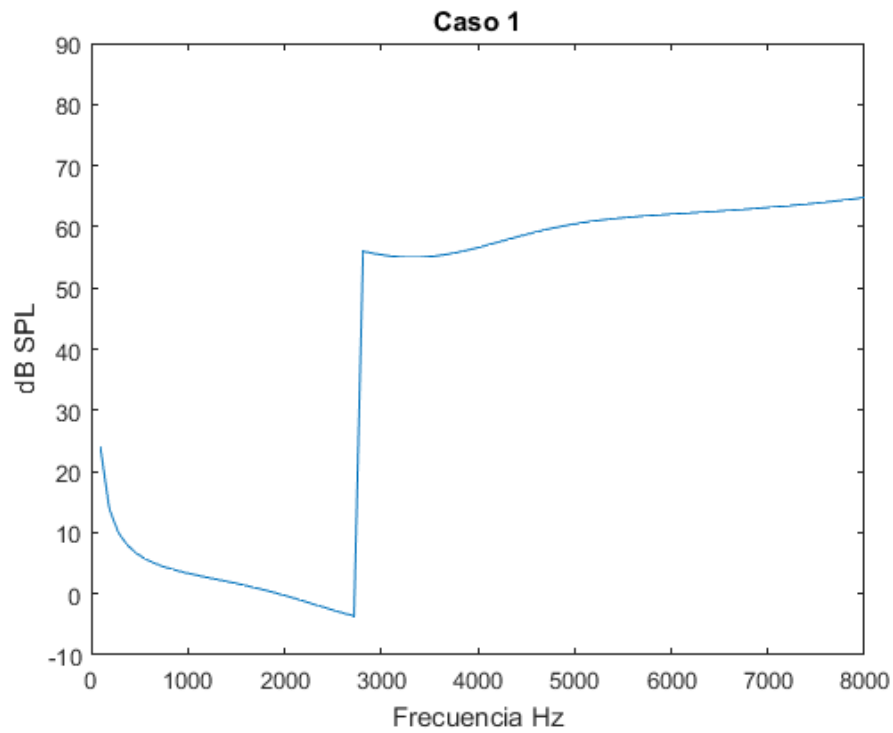


Figura 7 - Representación umbral de audición caso 1, corte 1000 Hz

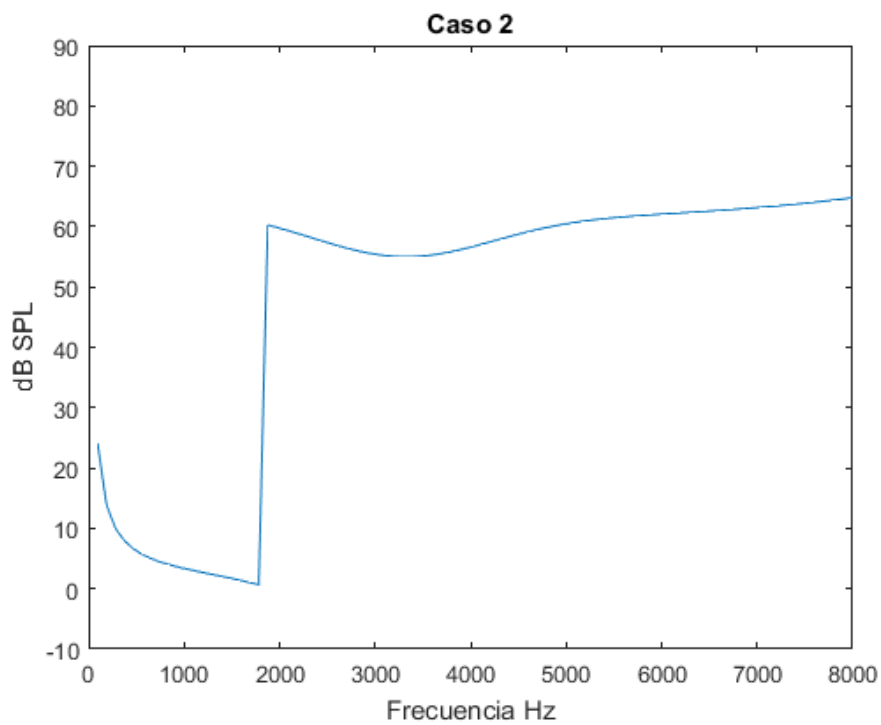


Figura 8- Representación umbral de audición caso 2, corte 500 Hz

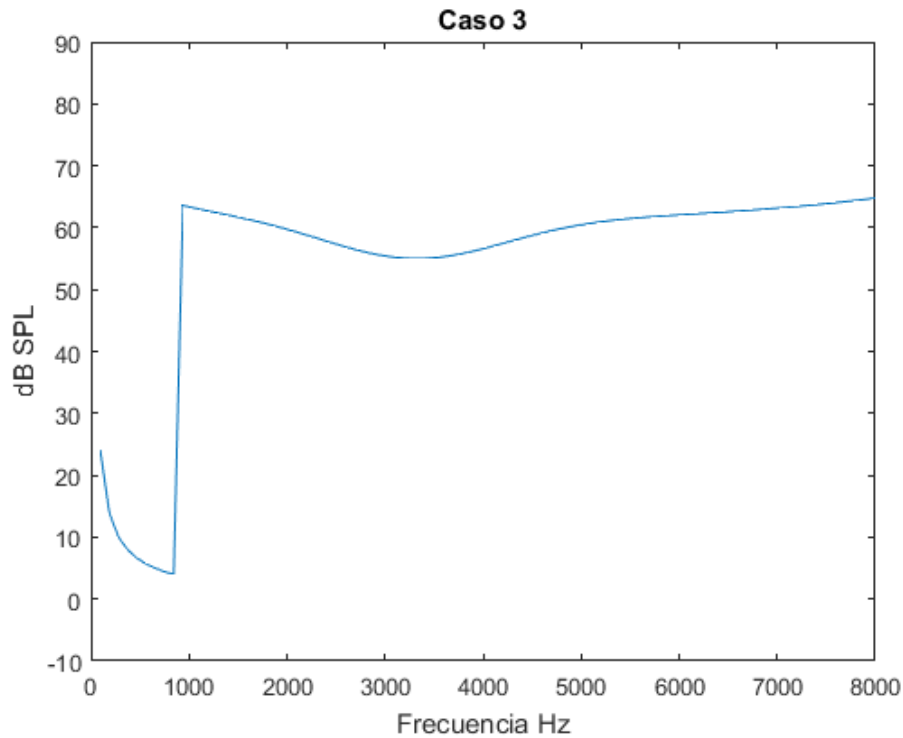


Figura 9- Representación umbral de audición caso 3, corte 312Hz

Todo lo anteriormente expuesto se presenta, en un diagrama de bloques del simulador en la figura 15, que está implementado en una función principal llamada *oído_solapado_or_v2_RIAI* y que sintéticamente es:

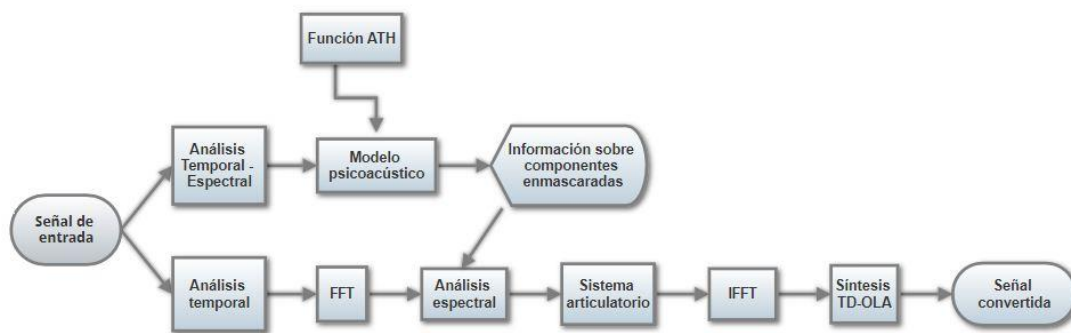


Figura 10 – Diagrama de bloques del simulador (A. Uriz)

La señal de entrada se descompone en segmentos de 32 ms de longitud, es decir, 512 muestras con una frecuencia de muestreo de 16 KHz, a continuación se le aplica la FFT, *Fast Fourier Transform* o Transformada rápida de Fourier, y, finalmente, se realiza la selección de la información espectral teniendo en cuenta la descartada mediante la función ATH introducida, correspondiendo al análisis

temporal-espectral del diagrama de bloques del simulador. A continuación, se debe realizar un ajuste de los segmentos para disminuir las discontinuidades en los extremos de los mismos, que corresponden al bloqueo de sistema articulatorio, y, posteriormente, se convierte, el espectro obtenido, al dominio del tiempo mediante la función IFFT, *Inverse Fast Fourier Transform*, o Transformada rápida de Fourier inversa. Por último, se sintetizan, mediante TD-OLA, *Time Domain Overlap – Add method*, o método de suma solapada en el dominio del tiempo, basado en la técnica fundamental de procesamiento digital de señales.

Por cada audio de entrada, se generan cuatro audios de salida, el primero de ellos es el original, que utilizaremos como base para las comparaciones, y los tres resultados restantes que corresponden con los audios que recogen las distintas pérdidas de audición. Todos ellos se analizarán a continuación. Se han analizados 200 ficheros de audio distintos, cada audio es una palabra de las listas utilizadas en este proyecto. Como resultado de la simulación, se han obtenido 800 registros de audio.

El apartado final del proceso se centra en la determinación objetiva de la inteligibilidad ayudándonos para ello de las medidas descritas anteriormente, PESQ, SNR y ALcons. Tratamos, conforme hemos indicado, cada uno de los audios y obtenemos, mediante el software y las funciones matemáticas descritas, la parte analítica de nuestro trabajo.

Como resumen de todo el proceso, en la figura 16, recogemos un diagrama completo del mismo.

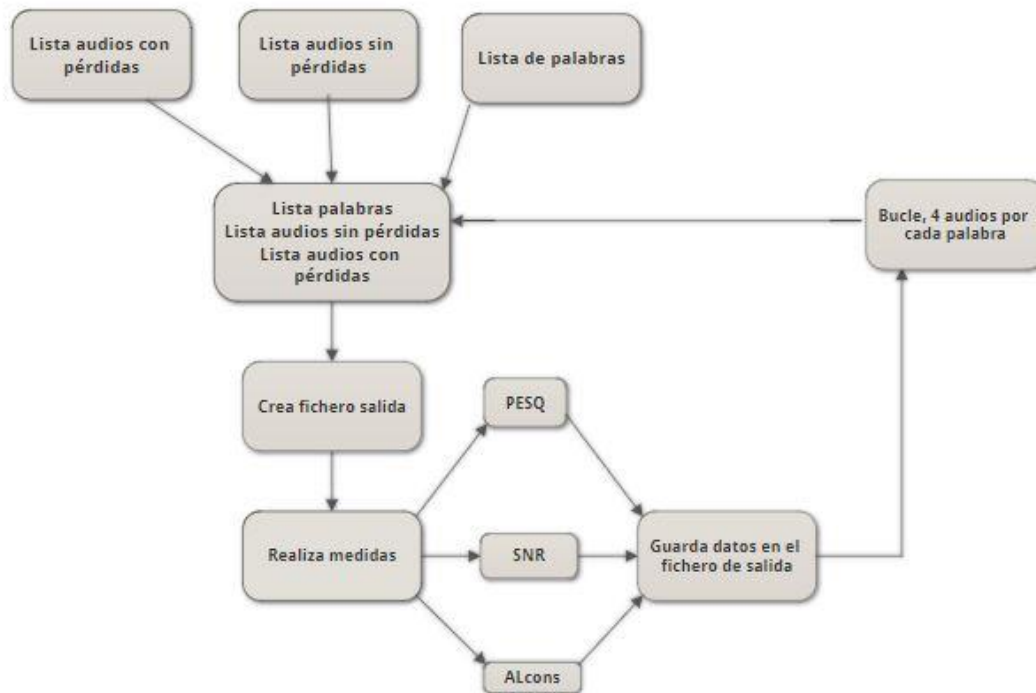


Figura 11 - - Diagrama de bloques del proceso completo

En el estudio analítico se han realizado tres medidas para cada audio simulado, comparando la señal original con las correspondientes a cada discapacidad auditiva simulada.

En primer lugar, se ha aplicado la medida PESQ, utilizando para ello la función establecida por Loizou, basada en el estándar P.862. Como valores de entrada, se ha introducido el audio original, los tres audios simulados, y la frecuencia de muestreo.

Esta función sólo soporta frecuencias de muestreo de 8 y 16 KHz, y por ello ha sido preciso modificar la frecuencia de muestreo de los audios obtenidos, para lo que se ha hecho uso de la función *resample(X, P, Q)*, esta sentencia remuestrea la secuencia en el vector X, que es la señal a modificar, P/Q veces la tasa de muestreo original. Por lo tanto, teniendo una frecuencia de muestreo original de 48 KHz, para pasar a 16 KHz se ha utilizado *resample(x, 1,3)*.

Para la medición de la relación señal a ruido, se ha utilizado la función *snr*, perteneciente a la Toolbox de procesamiento de señales, función que devuelve en decibelios el valor de relación señal a ruido de cada señal introducida.

Por último, para obtener de los datos de la medida ALcons se ha utilizado una función escrita por Jacob Donley, basada en STI, *Speech Transmission Index*, que hace uso de la relación matemática descrita anteriormente en el apartado ALcons.

Los resultados así obtenidos se presentan en el capítulo siguiente.

Capítulo 5

5 Estudio

5.1 Estudio analítico

Comentaremos ahora los resultados analíticos obtenidos con las herramientas de medida que hemos elegido, con el tratamiento dado a los audios con Matlab, y gracias al código y su metodología de uso.

Por cada audio, hemos realizado medidas de PESQ, SNR y Alcons obteniendo las gráficas que expondremos seguidamente.

Para calcular las medidas, comparamos dos señales, la primera es el audio sin tratar, y la segunda señal toma cuatro valores, la señal sin pérdida de audición, y las tres señales con las distintas pérdida de audición establecidas. En la figura 17, podemos ver la distribución de las medidas comentadas anteriormente para la simulación de la discapacidad 1, que recordemos, atenúa las frecuencias superiores a 1000 Hz.

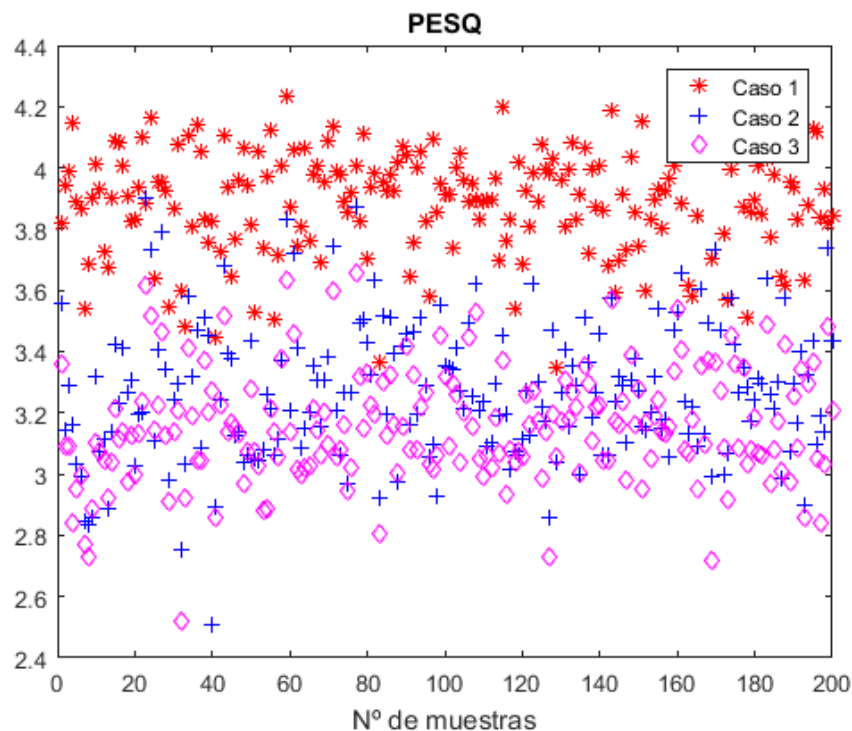


Figura 12 - Representación de resultados medida PESQ

En la representación, apreciamos cómo disminuye el valor de PESQ a medida que disminuye la calidad del audio simulado frente al original, y recordando que, el análisis de los resultados de la medida PESQ, se mueven entre 0 y 5, siendo el 5 una calidad excelente.

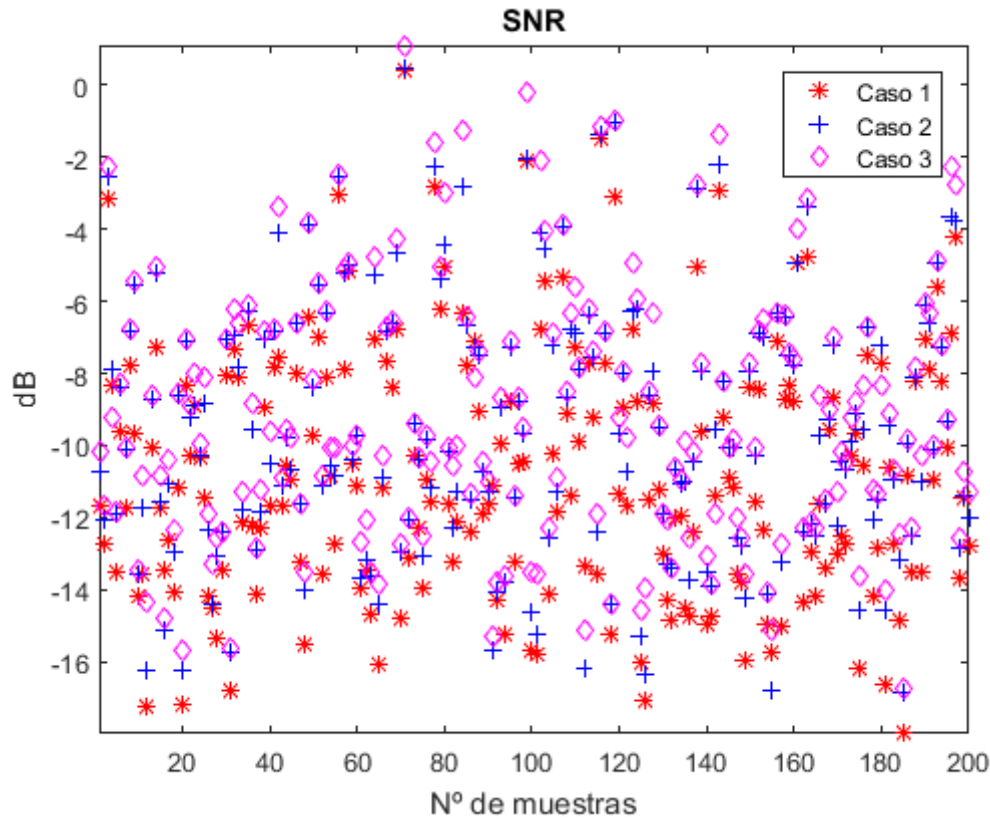


Figura 18 - Representación de resultados medida SNR

La medida SNR, que refleja la proporción existente entre la señal transmitida y el ruido que la corrompe en los audios estudiados, queda reflejada en la figura 18.

Apreciamos que la SNR aumenta a medida que empeora la discapacidad simulada, obteniendo aun así resultados muy similares para las distintas pérdidas de audición.

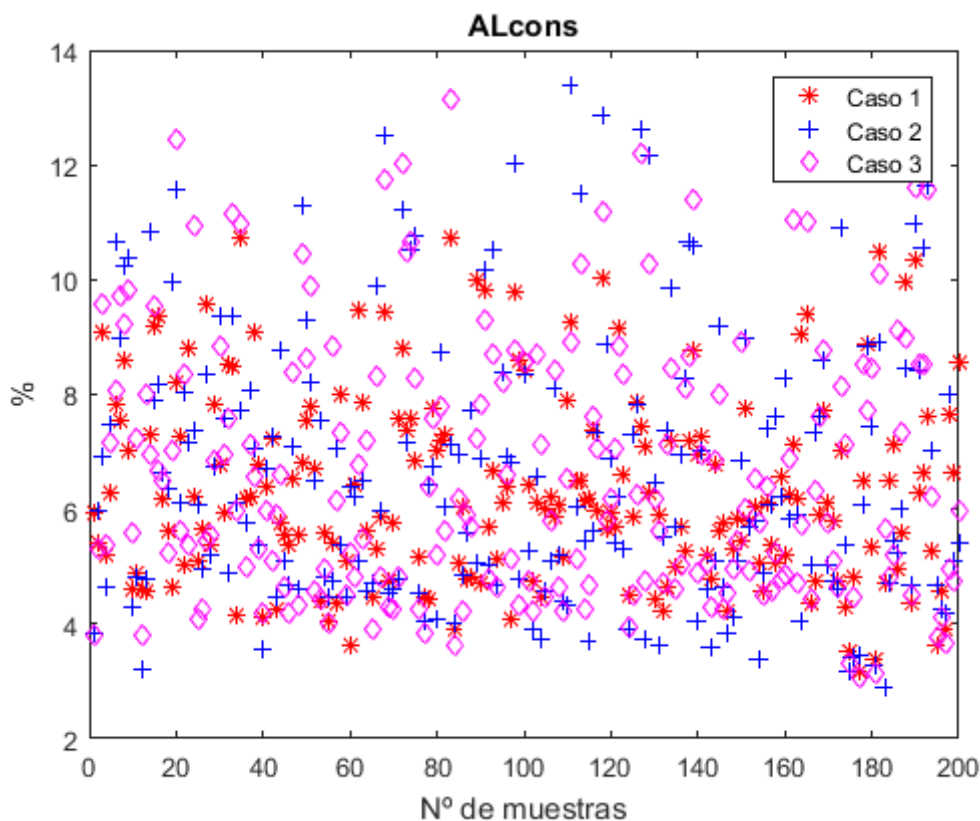


Figura 19 - Representación de resultados medida ALcons

Y, finalmente, en la figura 19, podemos ver el Alcons, la tercera de las mediciones previstas, que representa la pérdida porcentual de articulación de las consonantes en las 200 muestras estudiadas.

Inaceptable	Pobre	Aceptable	Buena	Muy buena
100-33%	33-15%	15-7%	7-3%	3-0%

Tabla 3 - Representación escala de la medida ALcons

En la tabla 3 recogemos una escala de cómo interpretar los datos. Vemos que, concretamente la discapacidad caso 1, se encuentra alrededor del 4-8%, mientras que las otras dos discapacidades muestran peores resultados.

El estudio de ALcons, analíticamente, tiende a ignorar factores externos, como ruido de fondo, distorsión o respuesta de frecuencia de la sala. Estos componentes pueden afectar considerablemente a la inteligibilidad de la palabra, ofreciendo así resultados optimistas. Por ello, en futuras actuaciones deberemos establecer situaciones más específicas, y así obtener resultados más ajustados.

A continuación, analizamos el fonema /k/ para mostrar un ejemplo del estudio. Este fonema se encuentra, por ejemplo en las palabras “cosa” y “casa”, y analizando “casa”, en la tabla 4 vemos los resultados obtenidos con nuestras herramientas.

PESQ	SNR	Alcon	Palabra	Tipo de discapacidad
4,00560900	-1,435784	3,38231600	casa	Discapacidad caso 1
3,30999000	-4.924552	3,26497400	casa	Discapacidad caso 2
3,06688100	-7,079888	3,14283800	casa	Discapacidad caso 3

Tabla 4 – Estudio analítico sobre la palabra ‘casa’

En la figura 20, vemos representado la detección de los formantes de la palabra ‘casa’. Una vez establecidos los formantes, se aprecia que el primer y segundo formante tienen unas frecuencias de 689 Hz y 1.680 Hz respectivamente.

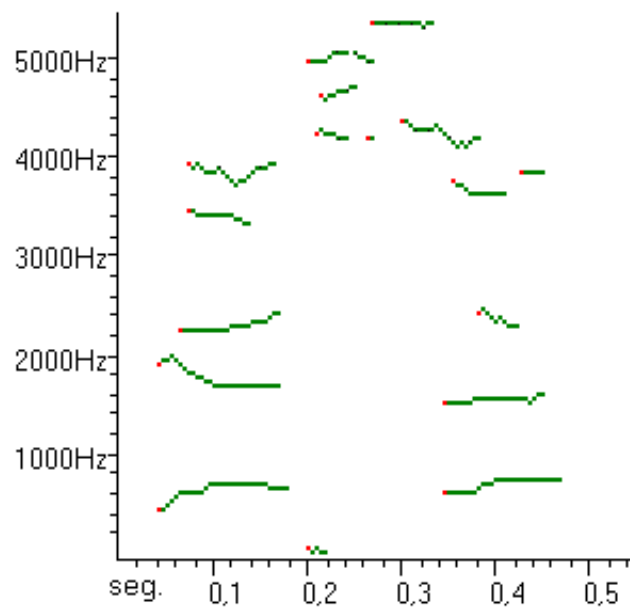


Figura 20 – Representación de los formantes de la palabra casa (Bernal, Bobadilla, & Gómez, 2000)

Podemos comprobar con los resultados de la PESQ que, a partir de la discapacidad 2, la que atenúa las frecuencias superiores a 500 Hz, los resultados empeoran drásticamente en comparación con la discapacidad 1. Esto se debe a que el primer formante se encuentra en la banda de 600 Hz, la cual no está atenuada en la discapacidad 1.

5.2 Estudio experimental

Para calcular, de manera objetiva, la inteligibilidad de los audios simulados, hemos hecho uso del *reconocimiento de voz online*. En unas condiciones adecuadas, y aisladas de sonidos fuertes exteriores, hemos utilizado un altavoz potente para reproducir todos los audios al mismo volumen, a la misma distancia del receptor, y hemos comprobado qué palabras ha sido capaz de reconocer sin ninguna distorsión, qué palabras ha cambiado de singular a plural o viceversa, en qué palabras la aplicación ha cambiado una letra y, por lo tanto, ha cambiado el significado de la palabra, qué palabras ha interpretado por otras con distintos significado, y cuándo no ha conseguido percibir nada. Para ello, hemos establecido una regla, asignando un número a cada posibilidad descrita anteriormente, conforme se indica en la tabla 5,

Caso	Número identificativo
Igual	5
Plural/Singular	4
Cambio de 1 consonante	3
Significado Diferente	2
Nada	1

Tabla 5 - Resumen de escala en estudio experimental

Y así, hemos obtenido, para cada tipo de discapacidad estudiado, las representaciones gráficas que mostramos a continuación.

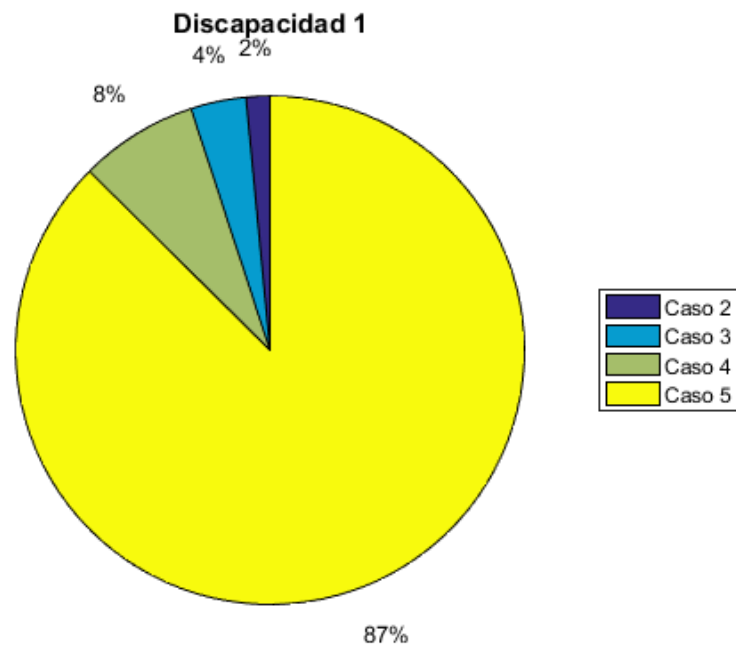


Figura 21 - Resultados experimento discapacidad caso 1

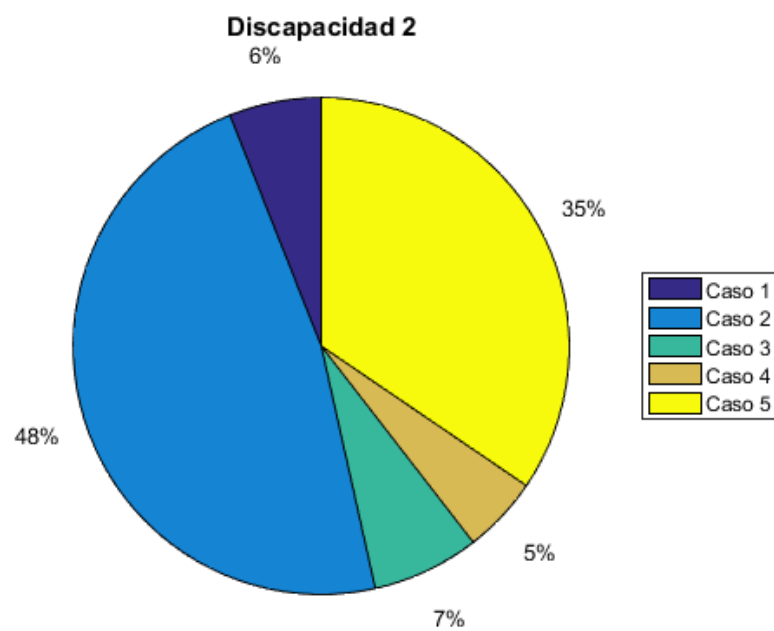


Figura 22 - Resultados experimento discapacidad caso 2

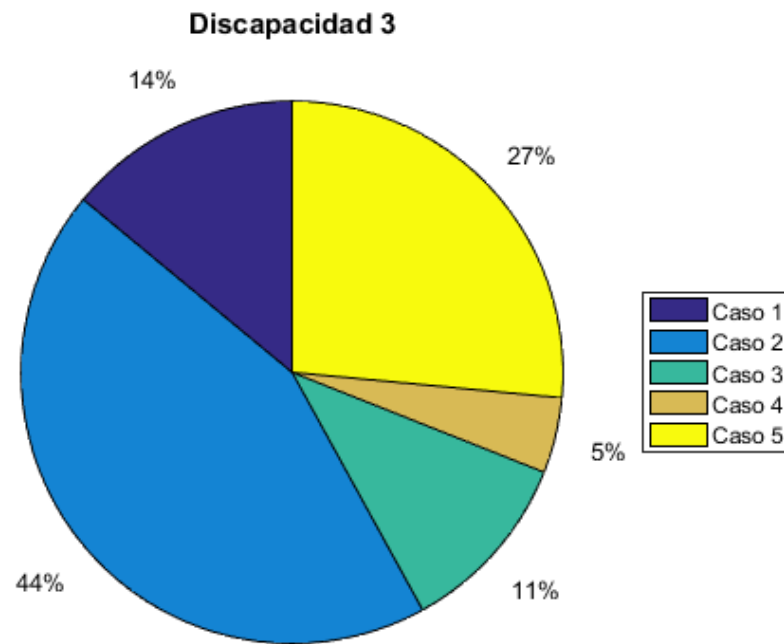


Figura 23 - Resultados experimento discapacidad caso 3

En las imágenes 21, 22 y 23 vemos, respectivamente, los resultados del método experimental. Podemos comprobar cómo, a medida que aumenta la discapacidad auditiva simulada, disminuye la inteligibilidad de las palabras. En el caso 1, se acierta el 87% de las palabras, es decir, la aplicación de reconocimiento de voz es capaz de identificar 174 de las 200 palabras. Mientras que en el caso 3, que es la discapacidad en que simulamos las pérdidas de audición desde 312 Hz, solo es capaz de identificar correctamente el 27% que equivale a 53 de las 200 palabras.

En el capítulo siguiente realizaremos un análisis de los resultados obtenidos en ambos experimentos.

Los datos resultantes de cada estudio se presentarán en un documento anexo.

Capítulo 6

6 Resultados

6.1 Análisis de los resultados

Nuestro proyecto puede ayudar al estudio de la inteligibilidad del habla mediante simulaciones de pérdidas de audición. Concluidos ambos estudios, analítico y experimental, podemos extraer varias conclusiones, analizando los resultados obtenidos, y que apoyan esta línea de investigación.

En primer lugar, la conclusión más importante que podemos extraer a partir de los análisis realizados, es la dificultad que sufren personas con discapacidades leves progresivas con la inteligibilidad de las palabras. En el estudio experimental se observa, que a medida que empeora la discapacidad simulada, los porcentajes de acierto disminuyen considerablemente, recordamos que para la discapacidad 1 el límite de frecuencia era de 1000 Hz, para la discapacidad 2 de 500 Hz, y para la discapacidad 3 de 312 Hz. Es de señalar que aunque, experimentalmente, la discapacidad 2 reconoce mal más palabras que la discapacidad 3, no obstante, es capaz de reducir el número de palabras no entendidas directamente.

Esto se puede traducir en una situación real, en que, una persona con pérdida de audición, encuentre problemas para entender lo que dice otra persona, y se vea incapaz de seguir una conversación. Además estos problemas podrían aumentar considerablemente si la conversación se produce en localizaciones con ruido de fondo, por ejemplo en una cafetería o en una calle. No obstante, hay que tener en cuenta, que cuando se trata de frases completas, el ser humano puede ser capaz de obtener el significado completo aunque no haya percibido todas las palabras.

En segundo lugar, el límite de rango de frecuencias que perciba el sujeto es muy importante a la hora de establecer la inteligibilidad, ya que, dependiendo de ese límite, se dejarán de percibir ciertos sonidos o llevará a la confusión de distintos fonemas.

Cabe destacar, igual que antes hemos comentado, que el ser humano puede predecir una palabra a través del contexto de una frase, puede hacer lo mismo con las consonantes frente a las vocales. Las consonantes, aunque posean menos energía, contienen mayor cantidad de información fonética que las vocales.

En la tabla 6 se recogen los resultados obtenidos en el estudio analítico de la palabra “dato”, en una representación completa de los archivos, donde se muestran los números reales obtenidos en el experimento analítico, de todos los audios originales, con sus respectivos audios simulados. Se observa que en el estudio del audio original, la PESQ es 4.5 en todas las muestras, siendo esa una calidad casi excelente disminuyendo conforme aumenta el grado de discapacidad. En cuanto a la relación señal a ruido, vemos cómo cambia conforme sea la calidad del audio y finalmente, contemplamos que el porcentaje de pérdida de articulación disminuye también con el grado de discapacidad.

Método	PESQ	SNR	ALcons	Palabra	Tipo de Discapacidad
clean :	4.500.000	0.000000	6.077.059	dato	salida_sano_18.wav
noisy :	3.904.711	-0.314133	5.615.118	dato	caso1_discapacidad_18.wav
noisy :	3.265.558	-1.849945	6.348.530	dato	caso2_discapacidad_18.wav
noisy :	2.976.412	-4.201681	5.248.535	dato	caso3_discapacidad_18.wav

Tabla 6 – Resultado del estudio analítico de la palabra ‘dato’

Podemos ver la conexión con el resultado experimental observando sus resultados para el mismo audio, reflejados en la tabla 7.

Apreciamos que en la discapacidad de caso 1, es capaz de identificar la palabra original, en el caso 2, el reconocedor de voz automático confunde con el plural de la palabra original. En cuanto a la discapacidad 3, vemos que al confundir sólo una letra de la palabra, el significado de ella cambia totalmente.

Palabra	Caso 1	Caso 2	Caso 3	Caso 1	Caso 2	Caso 3
dato	dato	datos	gato	5	4	3

Tabla 7 – Resultado del estudio experimental de la palabra ‘dato’

Comparando ambas tablas, 6 y 7, apreciamos que los resultados de ambos experimentos tienen simetría conjunta, pues en el caso 1, donde se escucha la palabra original, la PESQ tiene un valor de 4.5, considerado de buena calidad, la relación señal a ruido es alta, y el porcentaje de pérdida de articulación de las consonantes se encuentra al 5%, significando inteligibilidad muy buena.

6.2 Conclusiones

Los resultados obtenidos reflejan la concordancia entre ambos estudios, experimental y analítico, perdiéndose inteligibilidad a medida que aumenta el grado de la discapacidad.

Por tanto, las herramientas y la metodología empleadas, se han reconocido como válidas para el estudio planteado, si bien deben mejorarse para su optimizar sus aplicaciones futuras.

Para la realización de este proyecto se ha utilizado un reconocedor de voz automático en el estudio experimental, y aunque nos ofrece una buena calidad, se podrían mejorar los resultados obtenidos implementando un software con mejores prestaciones, y realizando a su vez las pruebas en ambientes más controlados, como salas específicas destinadas al estudio del audio.

En el experimento analítico, las medidas utilizadas nos han servido para establecer un rango de inteligibilidad orientativo, pero se debería tener en cuentas posibles mejoras, por ejemplo en las simulaciones de pérdidas de audición, pudiendo mejorar éstas perfeccionando el software utilizado. Otro aspecto a tener en cuenta, como establecimos antes, es que la medida ALcons tiende a ignorar factores externos produciendo resultados optimistas.

Debemos aclarar que las medidas de calidad usadas en el proyecto ofrecen buenos resultados en conjunto, pues se debe tener en cuenta que si analizáramos la inteligibilidad de los audios con sólo una medida, no tendríamos datos suficientes para establecer la inteligibilidad de los registros y establecer unos resultados válidos.

En definitiva, la inteligibilidad de las palabras varía sustancialmente, dependiendo de aspectos subjetivos, de los ambientes,..., pero, el camino aquí estudiado, se muestra como válido para, una vez establecida la casuística a tratar, con las herramientas elegidas, y otras complementarias que se incorporen, podamos establecer patrones, cada vez más completos y fiables, que posibiliten actuaciones para la mejora y la prevención de estas discapacidades.

6.3 Líneas futuras

En este proyecto se ha realizado un estudio sobre la inteligibilidad de las palabras, para distintas pérdidas de audición establecidas, ayudándonos, para ello, de unos audios, simulados y preparados para tal fin, a los que hemos sometido a análisis, merced al uso de un programa informático específico, y cuyos resultados hemos podido medir con unas herramientas elegidas para el caso. Demuestra que es una metodología viable para el estudio exhaustivo de las discapacidades auditivas de manera que, perfeccionando las herramientas de tratamiento y medición, la metodología puede aportar muchísimos datos que faciliten la prevención y el tratamiento de estas deficiencias.

El estudio sobre la inteligibilidad en las discapacidades auditivas humanas presenta, como hemos dicho, una muy amplia casuística, dado que, cada persona, puede presentar un distinto umbral de audición, y, para similares discapacidades, que el origen de éstas pueda ser muy diferente. Para un análisis exhaustivo del problema, deberían realizarse estudios, con una amplia muestra de personas, inmersos en diferentes situaciones y ambientes (una sala sin ruido, una cafetería, un concierto, etc) que permitan determinar y clasificar más específicamente las discapacidades auditivas para así poder establecer premisas concretas, y cada vez más realistas, en el desarrollo de las herramientas informáticas.

En el futuro, se debe estudiar la inteligibilidad en oraciones completas, pues, como hemos comentado antes, una persona con pérdidas de audición puede ser capaz de intuir una palabra concreta analizando el contexto y el posible significado de la oración completa; en la oración “Me duele la boca”, donde el individuo escucha “bota” en sustitución de “boca”, va a comprender que el locutor no dice “bota”,

pues la frase completa no tendría significado coherente. Debemos, pues, grabar una exhaustiva serie de oraciones, fonéticamente equilibradas, para poder proceder a realizar las simulaciones, e ir completando estudios específicos cada vez más exhaustivos.

El camino a seguir debe, en cualquier caso, mantener la comparación de los resultados teóricos, con los extraídos del estudio sobre las personas, con discapacidades auditivas parecidas a las simuladas, para poder estimar con estadísticas, cada vez más reales, la fiabilidad de los experimentos y su aproximación a la realidad a estudio.

Capítulo 7

7 Marco regulador

La normativa específica, que regula y ampara las discapacidades auditivas, es relativamente escasa y, en todo caso reciente, pues la sensibilidad social hacia esta problemática concreta es de una muy cercana actualidad.

La pérdida de audición debe suponer una discapacidad funcional del 33% de la persona afectada, para ser considerado como una discapacidad amparable según recoge la ley general de Derechos de las personas con discapacidad y de su inclusión social, recogida en Real Decreto Legislativo 1/2013. Esto no suele ocurrir pues, las discapacidades auditivas, no están suficientemente estudiadas y documentadas y, socialmente, no son, en la mayoría de los casos, reconocidas como tales.

El desarrollo legislativo sobre la *lengua de signos*, en España, íntimamente relacionada con el caso que nos ocupa, se caracteriza por estar atomizada y dispersa en el marco legal de cada uno de los territorios autonómicos. La primera ley a nivel nacional data del 1995, el Real Decreto 2060/1995, y en ella se establece, por primera vez, el título de Técnico Superior en Interpretación de la Lengua de Signos. Esta ley se ha ido perfeccionando hasta su última modificación recogida en la Ley 27/2007.

En relación directa con el tema que nos ocupa, encontramos la legislación que aborda los problemas del ruido y la contaminación acústica. Su primera regulación data del 1961, el Reglamento de Actividades Molestas, Insalubres, Nocivas y Peligrosas (BOE, 1961). Esta fue la primera norma que intentó regular las materias de su título, y entre ellas el ruido. Era una ley que atendía más a reglar las relaciones sociales y de convivencia, y no atendía, como ocurre con la nueva sensibilidad de hoy en día, los enfoques de cómo ayudar a mejorar el desarrollo de las personas discapacitadas o la prevención de esa problemática. Esta ley, más tarde, se fue completando con reglamentos y ordenanzas de las distintas administraciones, principalmente los ayuntamientos, pero que mantenían todavía

esos fines de regular principalmente la convivencia. Posteriormente, y ya con una creciente concienciación social sobre los problemas de las personas discapacitadas, estas reglamentaciones, comienzan a ser más exhaustivas, de manera que el ruido, como problema, se comienza a entender y analizar como el causante de múltiples distorsiones, y hasta de serias enfermedades. Hoy en día, e inmersos ya en esta nueva conciencia social, el marco legal queda establecido desde el mínimo recogido en el CTE (Código Técnico de la Edificación), y, para nuestro caso, en su documento DB HR, completado con las normativas particulares, desarrolladas sobre todo en los grandes núcleos de población, y, todo ello, dentro del marco fijado por la Europa a la que pertenecemos que, el 28 de julio del 2002, estableció la Directiva 2002/49/CE del Parlamento Europeo y del Consejo, cuyo objetivo es proporcionar una base para poder desarrollar y complementar un conjunto de estudios y medidas, a nivel comunitario, sobre el ruido emitido por distintas fuentes sonoras, vehículos, equipamiento industrial, etc., y desarrollar, además, las medidas correctoras a corto y medio plazo.

Si consideramos la contaminación acústica, concepto que engloba toda la problemática que nos ocupa, encontramos que, en la legislación española, sus referencias más antiguas datan del año 2003, habiendo sido actualizada en los años 2007 y 2009. El objetivo principal de esta regulación es establecer unas normas comunes y necesarias referentes al ruido con los objetivos de calidad y el control de las emisiones acústicas. La ley 37/2003 sólo era aplicable a determinados emisores acústicos. En este caso está especificada para actividades comerciales, industriales o de servicios y sus instalaciones, carreteras, ferrocarriles, puertos y aeropuertos, no contemplando otros posibles emisores acústicos más probables en la vida cotidiana como pueden ser ruidos vecinales, de las actividades cotidianas y de ocio. Esta norma, define la contaminación acústica como *“La presencia en el ambiente de ruidos y vibraciones, cualquiera que sea el emisor acústico que los origine, que implique molestia, riesgo o daño para las personas, para el desarrollo de sus actividades o para los bienes de cualquier naturaleza o que causen efectos negativos sobre el medio ambiente.”* Actualmente se establecen los límites recogidos en la figura 24 que diferencia por usos y tipo de estancias principalmente.

Tabla 2.1 Valores de aislamiento acústico a ruido aéreo, $D_{2m,nT,Air}$, en dBA, entre un recinto protegido y el exterior, en función del índice de ruido día, L_d .

L_d dBA	Uso del edificio			
	Residencial y hospitalario		Cultural, sanitario ⁽¹⁾ , docente y administrativo	
	Dormitorios	Estancias	Estancias	Aulas
$L_d \leq 60$	30	30	30	30
$60 < L_d \leq 65$	32	30	32	30
$65 < L_d \leq 70$	37	32	37	32
$70 < L_d \leq 75$	42	37	42	37
$L_d > 75$	47	42	47	42

⁽¹⁾ En edificios de uso no hospitalario, es decir, edificios de asistencia sanitaria de carácter ambulatorio, como despachos médicos, consultas, áreas destinadas al diagnóstico y tratamiento, etc.

Figura 24 – Resumen del límite de decibelios permitidos. (CTE, 2009)

Finalmente, AENOR, *Asociación Española de Normalización y Certificación*, ha elaborado normativas sobre acústica y vibraciones, AEN-CTN 074. El principal campo de actividad de este comité es la normalización en el campo de la acústica, en sus aspectos de terminología, métodos de medida de los fenómenos acústicos, su generación, transmisión y recepción, y todos los aspectos sobre sus efectos en el hombre y en el medio ambiente, excluyendo el campo de la electroacústica y las especificaciones de instrumentos utilizados para fines acústicos. Este comité ha elaborado distintas normas sobre los temas descritos anteriormente. La norma UNE-EN ISO 10140-1:2016 trata sobre el aislamiento acústico de los elementos de la construcción y las reglas de aplicación para productos específicos. (ISO, Acústica, 2016)

Como resultado de lo anterior, tenemos que la problemática asociada a las discapacidades auditivas es muy reciente de manera que las distintas administraciones tanto autonómicas como nacionales o europeas y organismos reguladores varios, están comenzando a desarrollar metodologías y medidas de control que limiten el ruido y mejoren la contaminación acústica para en definitiva conseguir un ambiente más saludable.

Hemos de señalar, finalmente, la LOPD, sobre protección de datos y la propiedad intelectual, pues el código utilizado para la simulación de pérdidas de audición nos fue amablemente cedido por Alejandro Uriz, propietario y autor intelectual del mismo como investigador del Laboratorio de Comunicaciones de la Universidad

Mar del Plata en Argentina Nuestro sincero agradecimiento pues sin su aportación no hubiera sido posible la realización de este trabajo.

Capítulo 8

8 Aspecto socioeconómico y planificación

Recogemos aquí una breve descripción de las distintas etapas del desarrollo de este proyecto, y se establece el presupuesto valorando los recursos utilizados., y señalamos cómo este trabajo, y sus líneas de actuación, pueden afectar de manera positiva a la sociedad.

8.1 Entorno socioeconómico

8.1.1 Entorno social

Ya hemos comentado que el reconocimiento de las discapacidades auditivas es muy reciente y, casi, está todo por hacer.

El camino fijado en este estudio es una aportación, no solo desde el punto de vista médico, al que estos estudios facilitarán, con la casuística creciente de bases de datos que aporten, la comprensión cada vez mayor de las discapacidades acústicas y, por tanto, su diagnóstico, sino también, para la sociedad en general pues, con mayor conocimiento de las causas y las consecuencias de estas discapacidades, en cuanto que afectan a la inteligibilidad, podrán adoptarse medidas de protección y prevención, cada vez más eficaces, que acomoden nuestro entorno para prevenir, o en todo caso mitigar, la aparición de discapacidades auditivas, y, todo ello, con el objetivo último de mejorar el confort humano, y preservar la calidad de la comunicación de las personas.

8.1.2 Entorno económico

El proyecto aquí desarrollado creemos que es el camino lógico para ahondar en el estudio de los problemas auditivos, y su profunda interrelación con la comunicación humana de una manera eficaz y económica.

La creación de múltiples y variadas bases de datos, simuladas con ordenador, es mucho más económico y factible que intentar establecerlas con el análisis de grandes grupos humanos donde la singularidad, las apreciaciones subjetivas, y un

sin fin de variables que afectan a cada individuo, hacen casi imposible establecer premisas homogéneas que nos hagan avanzar en la prevención de las discapacidades. Siempre será necesario contrastar los datos informáticamente obtenidos sobre las discapacidades teóricas, con la realidad de las personas, pero la casuística generada por aquellos, hará mucho más viable la tipificación y clasificación de las discapacidades reales.

Por tanto, y en un entorno económico cada vez más entrelazado, habrán de multiplicarse los estudios, cada vez más completos, afectando a disciplinas varias, que permitan predecir mejor la inteligibilidad asociada a las distintas discapacidades auditivas y atender así a mejorar la comunicación humana.

Se necesitarán inversiones para la mejora del software al objeto de poder ampliar la casuística que puedan manejar, preparando científicamente extensos registros de palabras y frases completas, generados en distintas localizaciones y ambientes, y con diferentes locuciones, que incorporen la variable acento, en el estudio de la inteligibilidad de las palabras por parte del oyente que presente una discapacidad auditiva.

8.2 Presupuesto

A continuación, vamos a presentar una valoración económica del proyecto. En primer lugar, en la tabla 8 detallaremos el coste del personal dedicado a este proyecto.

<u>Coste personal</u>				
Apellidos y nombre	Categoría	Horas	Hora/mes	Importe (€)
Peláez Moreno, Carmen	Ingeniero	40	25	1000
Redondo Martínez, Patricia	Estudiante	285	5,5	1567,5

Tabla 8 – Coste personal

Por otra parte, tenemos que tener en cuenta los materiales utilizados para el desarrollo del estudio que establecerán los gastos materiales del proyecto. En este caso, sólo se ha utilizado un ordenador con el que se han realizado las simulaciones y el presente documento, cuyo importe fijaremos en 750€ y unos altavoces inalámbricos valorados en 60€. Estos gastos se reflejan en la tabla 9.

<u>Coste material</u>	
Tipo de material	Importe (€)
Ordenador	750,00
Altavoces	60,00

Tabla 9 – Coste material

En la tabla 10 vemos los gastos tenidos en cuenta como gastos indirectos, no están relacionados con el objetivo principal del proyecto pero se producen para poder lograrlo.

<u>Coste general</u>	
Tipo	Importe (€)
Gasolina	50,00
Teléfono	30,00
Luz	100,00

Tabla 10 – Coste general

En la tabla 11, vemos que el importe total del presupuesto de este proyecto es de CUATRO MIL NOVENTA Y SIETE CON NOVENTA Y SIETE EUROS.

CONCEPTO	CANTIDAD	PRECIO	IMPORTE
Gasto material	1,00	810,00	810,00
Gasto personal	1,00	2.567,50	2.567,50
Otros gastos	0,50	185,00	9,25
		Suma	3.386,75
	0,00%	GG+BI ¹	0,00
	21,00%	IVA	711,22
		Importe TOTAL	4.097,97

Tabla 11 – Presupuesto total del proyecto

¹ Los gastos generales y el beneficio se presuponen como cero al ser un estudio académico.

8.3 Planificación

El desarrollo de este proyecto ha implicado varias fases de estudio y aprendizaje. Una vez determinados los objetivos, y afianzados los conocimientos, se procedió a la investigación sobre software y estudios relacionados con las discapacidades auditivas. A continuación se explican, con detenimiento, las diferentes fases del proyecto:

En primer lugar se realizó un estudio general para decidir la línea argumental del proyecto, y, qué era lo que realmente se quería estudiar. Se realizó un profundo análisis de las posibles líneas a seguir, y de los estudios que había publicados al respecto. Completándolos con varios cursos de Coursera: La música como biología, por la Universidad de Duke, y Procesamiento Digital de Audio para Aplicaciones Musicales por las Universidades Pompeu Fabra de Barcelona y Standford.

Una vez claros los objetivos, planteamos y resolvimos una fase de contextualización, en la que se realizaron análisis de los requerimientos necesarios, y de los procedimientos a desarrollar: ¿qué necesitamos?, y ¿cómo lo vamos a conseguir?. En el análisis de los requerimientos, consideramos los posibles lenguajes de programación que podíamos manejar, realizando para los procedimientos, una profunda investigación sobre estudios, ya desarrollados, que trataran sobre las discapacidades auditivas, y las simulaciones de éstas.

Para el desarrollo de la simulación, hicimos una exhaustiva investigación sobre simuladores ya existentes y trabajos funcionalmente similares. A results de ello, encontramos dos simuladores realmente capaces de proporcionarnos la ayuda que necesitábamos, y, así, contactamos con sus creadores, primero con Alejandro Uriz, investigador del Laboratorio de Comunicaciones de la Facultad de Ingeniería de Mar del Plata (Argentina), y, seguidamente, con Amir Rahimzadeh del Institute of Electronic Music and Acoustics (IEM) en Austria.

Realizamos un estudio completo sobre ambos simuladores, comentando con los autores los pros y los contras que apreciábamos para la resolución de nuestro trabajo, y que amablemente nos ayudaron a resolver. Finalmente escogimos el simulador de Alejandro Uriz, pues se ajustaba mejor a nuestra idea de proyecto.

Este simulador se implementa siguiendo el modelo de MPE- Audio Layer, estándar de comprensión de audio basado en un modelo auditivo humano.

Una vez estudiado el código, y con permiso del autor, fuimos haciendo las variaciones pertinentes para adecuarlo a las necesidades requeridas para el correcto desarrollo de nuestro proyecto.

Necesitábamos, a continuación, una serie de audios adecuados para poder tratarlos en nuestro simulador ya preparado. La búsqueda de estos audios llevó más tiempo del previsto pues queríamos una lista de audios en español, y debían de estar grabados en unas condiciones adecuadas. Por razones de autor y copyright, empresas dedicadas a la audiología y estudio de las discapacidades, no nos los podrían proporcionar. Seguimos buscando, y encontramos que, el Departamento de Filología de la Universidad Autónoma de Barcelona, había elaborado un Corpus escrito de palabras en español orientadas al estudio de la inteligibilidad y que, con ayuda de Widex Audifonos S.A., habían elaborado, en 1993, un CD con estas listas. Para conseguir este CD, fue preciso solicitar una copia a la Biblioteca Nacional de España.

Existen otras bases de datos que podrían habernos servido para realizar la simulación, como Albayzin, que está diseñada por un consorcio de seis grupos de investigación, y producida por la empresa Page Iberica S.A. y que consta de 3 corpus de señales, grabadas a 16 KHz, y producidas por 304 locutores de la variedad central del castellano, un corpus fonético, uno geográfico y un “Lombard”, una mezcla de los anteriores.

Finalmente, la Biblioteca Nacional atendió a nuestro requerimiento, y nos prestó el CD solicitado, teniendo acceso así a la base de audios de Juan María Garrido, que está especialmente destinada a pruebas de inteligibilidad, y que entendimos era la más adecuada al objeto de este estudio.

Obtenidos los audios, 10 listas, con 20 audios=palabras cada una, pudimos ejecutar la simulación de cada uno de ellos. Para el mejor tratamiento de los audios, con la ayuda de un programa llamado Audacity, que es un editor de audio gratuito, separamos cada palabra en un archivo diferente.

Después de hacer las simulaciones, procedimos a su estudio, en dos fases; una primera, de forma matemática, donde analizamos distintas medidas de la calidad de audio, y una segunda fase, de manera experimental, donde realizamos una escucha de los audios tratados y, con un reconocimiento de voz, estudiamos cómo de bien se entienden las palabras con cada grado discapacidad establecido.

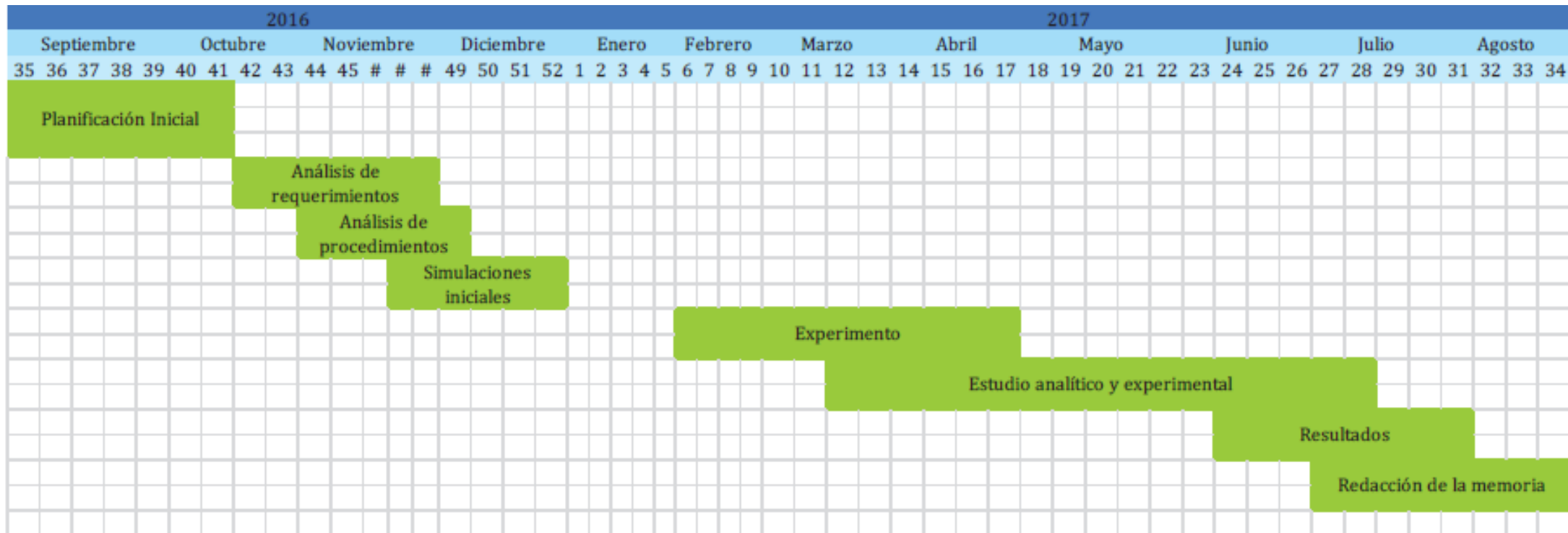
Todo este proceso nos aporta unos resultados que, analizados y comparados, teniendo presentes las limitaciones reconocidas, nos permiten establecer unas conclusiones sobre la inteligibilidad y marcar futuras actuaciones para ahondar en este camino.

El desarrollo de todo el proyecto en el tiempo queda recogido en la tabla 12, donde se recoge el cómputo horario empleado para el desarrollo de las distintas actividades. Y a continuación, podemos ver representada las tareas realizadas mediante el Diagrama de Gantt.

Tarea	Horas (Aprox)
Planificación inicial	15
Análisis de requerimientos	25
Análisis de procedimientos	30
Simulaciones iniciales	35
Experimento	60
Estudio matemático	30
Estudio experimental	30
Resultados	40
Redacción de la memoria	20
Total	285

Tabla 12 – Resumen de tareas realizadas

Diagrama de Gantt



Capítulo 9

9 Bibliografía

- A. Uriz, P. A. (s.f.). *Simulador de Deficiencias Auditivo implementado en Matlab*. Mar del Plata, Argentina.
- Baer, B. M. (2002). Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. *Journal of the Acoustical Society of America*, 112(3), 1133-1144.
- Ben Gold, N. M. (1999). *Speech and audio signal processing*. United States of America: John Wiley & Sons.
- Bernal, J., Bobadilla, J., & Gómez, P. (2000). *Reconocimiento de voz y fonética acústica*. Madrid: RA-MA.
- BOE. (1961). *Reglamento de Actividades Molestas, Insalubres, Nocivas y Peligrosas*. Madrid.
- Borja, A. (2011). *Neuropsicología Audioverbal*. Lima, Perú.
- Bradley, J., & Sato, H. (1986). *The intelligibility of speech in elementary school classrooms*.
- Bradlow, A. R. (1994). *A comparative acoustic study of English and Spanish vowels*. Nueva York.
- Canellada, M., & Madsen, J. (1987). *Pronunciación del español. Lengua hablada y literaria*. Madrid: Castalia.
- Chaix, B. (2016). *Audiograma vocal*. Obtenido de <http://www.cochlea.eu>: <http://www.cochlea.eu>
- Clark, J. (1981). *Uses and abuses of hearing loss classification*. ASHA.
- Crocker, M. J. (1997). *Encyclopedia of Acoustics*. John Wiley & Sons.
- CTE. (2009). *DB HR, Protección frente al ruido*.
- Curso de acústica creado por GA*. (2003). Obtenido de Universidad del País Vasco: <http://www.ehu.eus/acustica>
- Donley, J. (2015). *ALcons*. Wollongong, Australia.

- Embedded Acoustics*. (s.f.). Obtenido de Speech Transmission Index:
<http://www.embeddedacoustics.com>
- Garay, J. (2013). *Entendimiento de la relación señal a ruido y la atenuación*. Obtenido de Nireleku: <https://www.nireleku.com/>
- García, A. F. (2016). *Movimiento ondulatorio*. Obtenido de <http://www.sc.ehu.es/>
- ISO. (1993). International Organization of Standardization. *11172-3. Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio*.
- ISO. (1994). International Organization of Standardization. *Generic Coding of Moving Pictures and Associated Audio*. Ginebra, Suiza.
- ISO. (2016). Acústica. *Medición en laboratorio del aislamiento acústico de los elementos de construcción*.
- J. Caluppon, H. F. (2000). Simulation of hearing impairment based on the Fourier Time Transform. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 857-860.
- Jorge Sommerhof, C. R. (s.f.). *Evaluación de la inteligibilidad del habla en español*. Chile. Obtenido de Universidad de Chile.
- Juilland Alphonse, E. C.-R. (1964). *Frequency dictionary of spanish words*. The Hague.
- Leo L. Beranek, T. M. (2012). *Acoustics: Sound Fields and Transducers*. Oxford: Elsevier. doi:10.1016/B978-0-12-391421-3
- Loizou, P. C. (2007). *Speech Enhancement*. Taylor and Francis.
- Miyara, F. (2017). El ruido y la inteligibilidad de la palabra. Obtenido de Facultad de Ciencias exactas, ingeniería y agrimensura.
- National Institute on Deafness and Other Communication Disorders (NIDCD). (Marzo de 2017). Obtenido de Pérdida de audición inducida por el ruido: <https://www.nidcd.nih.gov/es>
- Oppenheim, A. (1999). *Discrete-time signal processing*.
- P.862, R. U.-T. (2001). *Evaluación de la calidad vocal por percepción*.
- Perez, C., & Zamanillo, J. (2003). *Fundamentos de la televisión analógica y digital*. Cantabria: Universidad de Cantabria.

- R. Arelhi, D. (2003). A MATLAB Simulink Implementation of Speech Masking Based on the MPEG Psychoacoustic Model I. *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*, 543-548.
- Rajadurai, J. (2007). Intelligibility studies: a consideration of empirical and ideological issues. *World Englishes*, 87-98.
- Slaney, M. (1998). *Auditory Toolbox*.
- Slaney, M. (1998). *Auditory Toolbox*. Obtenido de <https://engineering.purdue.edu>
- T. Painter, A. S. (2000). Perceptual Coding of Digital Audio. *Proceedings of the IEEE*, 88, 451-513.
- UAB, D. d. (Marzo de 2017). *Audiometría verbal*. Obtenido de http://liceu.uab.cat/~joaquim/applied_linguistics/speech_pathology/audio_metría/audiometria.html#La_audiometr_a_verbal
- Widex Audífonos (1993). *Audiometría verbal*.